The Open University

M303

Further pure mathematics

Book D

Metric spaces 1

This publication forms part of the Open University module M303 *Further pure mathematics*. Details of this and other Open University modules can be obtained from Student Recruitment, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)300 303 5303; email general-enquiries@open.ac.uk).

Alternatively, you may visit the Open University website at www.open.ac.uk where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2014. Second edition 2016.

Copyright © 2014, 2016 The Open University

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, transmitted or utilised in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without written permission from the publisher or a licence from the Copyright Licensing Agency Ltd. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS (website www.cla.co.uk).

Open University materials may also be made available in electronic formats for use by students of the University. All rights, including copyright and related rights and database rights, in electronic materials and their contents are owned by or licensed to The Open University, or otherwise used by The Open University as permitted by applicable law.

In using electronic materials and their contents you agree that your use will be solely for the purposes of following an Open University course of study or otherwise as licensed by The Open University or its assigns.

Except as permitted above you undertake not to copy, store in any medium (including electronic storage or use in a website), distribute, transmit or retransmit, broadcast, modify or show in public such electronic materials in whole or in part without the prior written consent of The Open University or in accordance with the Copyright, Designs and Patents Act 1988.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed in the United Kingdom by Halstan & Co. Ltd, Amersham, Bucks.

ISBN 978 1 4730 2039 9

2.1

Contents

Chapter 13 Distance and continuity **1**

Introduction **3**

1 Functions and sequences **4**
1.1 The modulus function 6
1.2 Sequences 11

2 Continuity on the real line **23**
2.1 Continuous and discontinuous functions 23
2.2 Rules for continuous functions 29
2.3 Continuity theorems 32

3 Continuity on the plane **36**
3.1 Functions from $\mathbb{R}^2$ to $\mathbb{R}$ 36
3.2 Sequences in $\mathbb{R}^2$ 37
3.3 Continuity on the plane 42
3.4 Properties of distance 49

Learning outcomes **51**

Solutions to exercises **52**

Chapter 14 Metric spaces and continuity 1 **63**

Introduction **65**

1 Continuity of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ **66**
1.1 Functions and distance 67
1.2 Properties of the Euclidean distance in $\mathbb{R}^n$ 68
1.3 Sequences in $\mathbb{R}^n$ 71
1.4 Continuity for functions between Euclidean spaces 75
1.5 Proof of the Triangle Inequality 83

2 Introducing metric spaces **87**
2.1 The definition of a metric space 87
2.2 Examples of metric spaces 89
2.3 Understanding the geometry of metric spaces 94

3 Sequences in metric spaces **98**

4 The definition of continuity in metric spaces **101**

Learning outcomes **104**

Solutions to exercises **105**

Chapter 15 Metric spaces and continuity 2 **113**

Introduction **115**

1 New metrics from old **115**
1.1 Transferring a metric 115
1.2 Metric subspaces 118
1.3 Product metrics 121
1.4 n-fold products 124

2 The Cantor metric **126**

3 Equivalent metrics **132**
3.1 Lipschitz functions 132
3.2 Equivalence of metrics 135
3.3 Convergence in product spaces 142

4 Spaces of functions **143**
4.1 Pointwise convergence 144
4.2 Uniform convergence 146

5 The max metric on $C[0, 1]$ **152**
5.1 Sequences in $C[0, 1]$ 161

Learning outcomes **164**

Solutions to exercises **165**

Chapter 16 Open and closed sets **177**

Introduction **179**

1 Closed sets **180**
1.1 Examples of closed sets 181
1.2 The algebra of closed sets 189

2 Open sets **194**
2.1 Examples of open sets 195
2.2 The algebra of open sets 201

3 Closure **207**
3.1 Elementary properties of closures 208
3.2 Further properties of closures 210
3.3 An alternative way of characterising closure 211
3.4 Closed sets and subspaces 215

4 Interior **216**

5 Boundaries of sets **220**
5.1 The algebra of boundaries 223

6 The size of sets **224**
6.1 Dense and nowhere dense sets 224
6.2 Countable and uncountable sets 227

7 Continuity and open and closed sets **231**

Learning outcomes **235**

Solutions to exercises **236**

Index **249**

Chapter 13

Distance and continuity

Introduction

You already know how useful and important it is to be able to determine the distance between points in the plane (or space); here you will see whether it is possible to define a useful notion of distance between more abstract objects such as functions, or elements from a group. In this book (and continuing in Book F), we develop a theory that enables us to talk about the distance between elements from a diverse range of different types of object (such as groups and function spaces) by introducing the theory of *metric spaces*: sets that have been equipped with some suitable method for determining a 'distance' between any two elements.

Although (as you will see in the next chapter) the definition of a metric space is a broad one, it gives rise to a rich theory that generalises and extends many ideas from (real) analysis with which you will already be familiar. In particular, it enables us to talk about the convergence (or otherwise) of sequences of objects other than real numbers, such as functions or sets. This in turn provides a way to extend the concept of continuity far beyond that of functions on the real line, ultimately enabling us to confirm that problems such as the following have (unique) solutions:

> Is there a continuous function $f\colon [0,1] \to [0,1]$ such that $f(0) = 0$ and, for $0 < s < 1$, f satisfies the differential equation
>
> $$\frac{df}{ds} = \tfrac{1}{2}\cos f(s)?$$

We will develop a method for showing that problems like this have a solution in Book F, Chapter 23, *Completeness*. Note that Wolfram Alpha gives the solution of the ordinary differential equation to be $f(s) = 2\tan^{-1}(\tanh(s/4))$.

This first chapter has a different character from that of the subsequent chapters in this book, in that much of it revises concepts from real analysis with which you should already be familiar from your previous studies. It begins with a review of sequences and functions of real numbers, followed by a reminder of the definition of continuity for functions on the real numbers.

This raises a question: how might we extend the notion of continuity to functions between more general sets? For example, which functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ should we describe as 'continuous'? As a first step towards answering this question, in Section 3 we generalise the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ to functions from $\mathbb{R}^2$ to $\mathbb{R}$. A crucial aspect of this process involves uncovering certain properties embedded in our idea of the 'distance between points'. In the next chapter, a further exploration of these concepts will lead to definitions of 'distance' and 'continuity' applicable not only to $\mathbb{R}^n$, but also in a much more general setting.

Chapter guide

Most of the material in Sections 1 and 2 should already be familiar to you, so it is likely that you will not need to spend that much time working through these sections. You might find it helpful to use the exercises to identify the parts that you need to concentrate on.

In Section 3 we start to look at a new situation and examine how our understanding of continuity for functions on the real line can be used to motivate a definition of continuity for functions mapping from the plane to the real line; this is our first tentative step towards metric spaces.

1 Functions and sequences

We start our investigation of continuity by clarifying some of the important terminology and notions on which the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ is based.

As in previous books of M303, we assume that you are familiar with the usual arithmetic operations of addition, subtraction, multiplication and division of real numbers; the different sorts of interval on the real line; manipulating simple inequalities involving real numbers; and the trigonometric and exponential functions.

A brief summary of the concepts with which you should be familiar is given in the first part of the *Handbook*.

We also assume that you are already familiar with the basic ideas associated with sets and subsets, such as membership, union and intersection. Many of these ideas were used in Book B.

As a reminder, we list some frequently seen sets of numbers, most of which you have already encountered in Book A and elsewhere in the module:

- the **natural numbers** $\mathbb{N} = \{1, 2, 3, \ldots\}$
- the **integers** $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$
- the **rational numbers** $\mathbb{Q} = \{p/q \colon p \in \mathbb{Z},\ q \in \mathbb{N}\}$
- the **real numbers** $\mathbb{R}$, which can be represented by infinite decimals
- the **irrational numbers** $\mathbb{R} - \mathbb{Q}$, which are those real numbers that are not rational, such as $\sqrt{2}$, π and e
- intervals of real numbers such as $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$, the **closed interval** with endpoints a and b, and $(a, b) = \{x \in \mathbb{R} : a < x < b\}$, the **open interval** with endpoints a and b.

In some texts the open interval is denoted by $]a, b[$.

Although in Sections 1 and 2 of this chapter we are only concerned with functions from $\mathbb{R}$ to $\mathbb{R}$, we record here the general definition of a *function* between two *arbitrary* sets. This has the advantage that we can use the same definition each time we change to a new setting.

Functions are also called *mappings* or *maps*. In some texts, the words 'mapping' and 'map' denote only continuous functions, but this is not our usage.

Definition 1.1 *Function between two arbitrary sets*

A **function** f is defined by specifying:

- a set A, the **domain** of f
- a set B, the **codomain** of f
- a **rule** $x \mapsto f(x)$ that associates with each element $x \in A$ a *unique* element $f(x) \in B$.

These are illustrated in Figure 1.1. The element $f(x)$ is called the **image** of x under f.

For $C \subseteq A$, the set $f(C) = \{f(x) : x \in C\} \subseteq B$ is called the **image set** of C under f, and for $D \subseteq B$, the set $f^{-1}(D) = \{x \in A : f(x) \in D\}$ is called the **preimage set** of D under f.

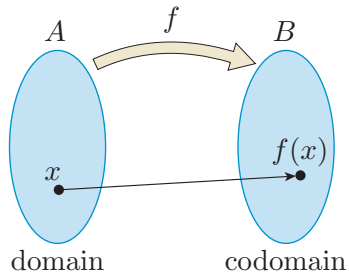


Figure 1.1 Definition of a function

There is some ambiguity in this notation in that f^{-1} does not necessarily denote the inverse function, which may not be defined, but it is used almost universally and one soon gets used to it.

Remarks

1. The preimage is also sometimes referred to as the inverse image.
2. When specifying functions symbolically we use one of three forms. For example, for the function from $\mathbb{R}$ to $\mathbb{R}$ given by $f(x) = x^2$, we use one of the following:

   $$f : \mathbb{R} \to \mathbb{R}, \quad f : x \mapsto x^2,$$
   $$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2,$$
   $$f(x) = x^2 \quad (x \in \mathbb{R}).$$

   In the first case we may sometimes simply write $x \mapsto x^2$ rather than $f : x \mapsto x^2$.

   Although it is good practice to include both domain and codomain when describing functions, they are occasionally omitted where the context makes them clear. This is most likely for real-valued functions of real numbers. In this case, the domain should be interpreted as the set on which the function is defined. Thus the domain of the function $f : x \mapsto \frac{1}{x}$ is $\mathbb{R} - \{0\} = \{x \in \mathbb{R} : x \neq 0\}$, since this function is defined for all non-zero real numbers.
3. In general the codomain of a function need not be the same as the image set of the domain. For example, if $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = x^2$, then the codomain of f is $\mathbb{R}$ but the image set of $\mathbb{R}$ under f is $f(\mathbb{R}) = [0, \infty)$, the non-negative real numbers.

Exercise 1.1

Which of the following rules define functions with domain $\mathbb{Z}$ and codomain $\mathbb{N}$?

(a) Square each integer and add 1.

(b) Associate with each integer the set of all its factors.

(c) Associate with each integer the number of its factors.

Exercise 1.2

For each of those rules in Exercise 1.1 that defines a function, determine the image set of $\mathbb{N}$ and the preimage set of $\{5\}$.

1.1 The modulus function

This subsection discusses the *modulus function* – a function that will prove vital in our study of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ and which possesses some of the key properties of a distance function.

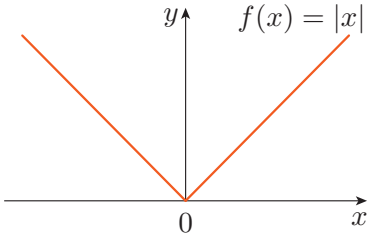


Figure 1.2 The modulus function

You may recall from Book A, Chapter 1 that the **absolute value** or **modulus** $|x|$ of a real number x is defined by

$$|x| = \begin{cases} +x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

For example, $|3| = 3$ and $|-3| = 3$. The graph of the modulus function is shown in Figure 1.2.

For each real number a,

$$0 \leq |a|, \quad a \leq |a|, \quad -|a| \leq a \quad \text{and} \quad |-a| = |a|. \tag{1}$$

For example, when $a = -2$,

$$0 \leq 2 = |-2|, \quad -2 \leq 2 = |-2|, \quad -|-2| = -2 \leq -2, \quad |-(-2)| = 2 = |-2|.$$

Note also that $|a| = 0$ if, and only if, $a = 0$.

Geometrically, $|a|$ gives the *distance*, without regard to direction, along the real line from a to 0.

For any real numbers a and b,

$$|ab| = |a|\,|b|, \tag{2}$$

$$|a|^2 = a^2. \tag{3}$$

For example, when $a = -2$ and $b = 3$,

$$|-2 \times 3| = |-6| = 6 = 2 \times 3 = |-2| \times |3|,$$
$$|-2|^2 = 2^2 = 4 = (-2)^2.$$

One of our main uses of the modulus function will be in *inequalities*, so we briefly discuss the solution set of the inequality

$$|x| < c,$$

where c is a constant in $\mathbb{R}$. The solution set depends on whether $c > 0$ or $c \leq 0$. For example, for $c = 2$, the set of x for which $|x| < c$ is the *open interval* $(-2, 2) = \{x\colon -2 < x < 2\} = \{x\colon |x| < 2\}$, as Figure 1.3 illustrates. However, for $c = -2$, there is no number x for which $|x| < c$, and so the solution set is the empty set, denoted $\emptyset$. In general, for $c > 0$ the solution set is the open interval $(-c, c) = \{x : -c < x < c\}$ and for $c \leq 0$ the solution set is the empty set $\emptyset$.

By *solution set* we mean the set of x for which the inequality holds.

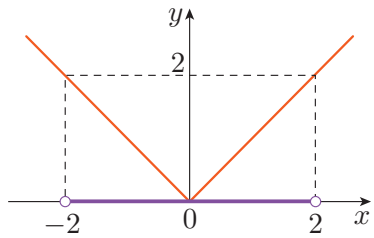


Figure 1.3 Solution set of $|x| < 2$

Thus for $c > 0$, the condition that

$$|x| < c$$

is equivalent to the double-sided inequality

$$-c < x < c.$$

This says, informally, that x lies between $-c$ and $+c$ on the real line: in other words, x is strictly less than a distance c from 0, as Figure 1.4 illustrates. The condition $-c < x < c$ is equivalent to saying that x satisfies the two simultaneous inequalities

$$-c < x \quad \text{and} \quad x < c.$$

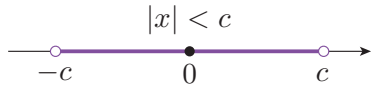


Figure 1.4 Real line showing $|x| < c$

We have similar results when the inequality is

$$|x| \leq c.$$

For $c \geq 0$, the solution set is the *closed interval* $[-c, c] = \{x\colon -c \leq x \leq c\}$. For $c < 0$, the solution set is the empty set $\emptyset$.

When $c = 0$, the solution set is $[0, 0]$, which is simply the one-point set $\{0\}$.

For $c \geq 0$, the inequality $|x| \leq c$ is equivalent both to the double-sided inequality $-c \leq x \leq c$ and to the pair of simultaneous inequalities $-c \leq x$ and $x \leq c$.

These results are summarised in the following proposition.

Proposition 1.2 *Solution sets of inequalities 1*

- For $c > 0$, the following inequalities are equivalent:
  (a) $|x| < c$, (b) $-c < x < c$, (c) $-c < x$ and $x < c$.
  Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R}\colon |x| < c\} = \begin{cases} (-c, c), & \text{if } c > 0, \\ \emptyset, & \text{if } c \leq 0. \end{cases}$$
- For $c \geq 0$, the following inequalities are equivalent:
  (a) $|x| \leq c$, (b) $-c \leq x \leq c$, (c) $-c \leq x$ and $x \leq c$.
  Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R} : |x| \leq c\} = \begin{cases} [-c, c], & \text{if } c \geq 0, \\ \emptyset, & \text{if } c < 0. \end{cases}$$

$[-0, 0] = \{0\}$

Exercise 1.3

Find the solution set of each of the following inequalities:

(a) $|x| < 3$; (b) $|x| \leq -3$; (c) $|3x| < 3$.

Worked Exercise 1.3

Prove that, for real numbers a and b,

$$|a| \leq |b| \quad \text{if, and only if,} \quad -|b| \leq a \leq |b|.$$

Solution

For $c \geq 0$, we have the equivalent expressions

$$|x| \leq c \quad \text{if, and only if,} \quad -c \leq x \leq c.$$

Setting $x = a$ and $c = |b|$, we obtain

$$|a| \leq |b| \quad \text{if, and only if,} \quad -|b| \leq a \leq |b|.$$

Exercise 1.4

Prove that, for real numbers a and b,

$$|a| \leq |b| \quad \text{if, and only if,} \quad -|b| \leq |a| \leq |b|.$$

We now consider more complicated inequalities.

Worked Exercise 1.4

For real numbers b and c, with $c > 0$, find the set of all real numbers x for which $|x - b| < c$.

Solution

The condition $|x - b| < c$ is equivalent to

$$-c < x - b < c,$$

which is the case if, and only if,

$$b - c < x < b + c.$$

Thus

$$\{x : |x - b| < c\} = (b - c, b + c).$$

So the required set is the open interval centred on b consisting of those points within a distance c of b, as shown in Figure 1.5.

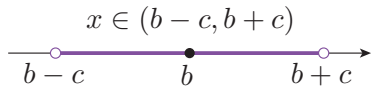


Figure 1.5 Real line showing $|x - b| < c$

A similar result holds for the inequality $|x - b| \leq c$, and so we have the following general results.

Proposition 1.5 *Solution sets of inequalities 2*

- For $b, c \in \mathbb{R}$ with $c > 0$,
$$\{x \in \mathbb{R} : |x - b| < c\} = (b - c, b + c).$$
- For $b, c \in \mathbb{R}$ with $c \geq 0$,
$$\{x \in \mathbb{R} : |x - b| \leq c\} = [b - c, b + c].$$

Exercise 1.5

Find the set of all numbers x for which:

(a) $|x - 1| < 3$; (b) $|x - 1| < 3$ and $|x + 1| < 3$.

In general, for any real numbers a and b,

$$|b - a| = |a - b|, \tag{4}$$

and this quantity gives the *distance* between a and b along the real line, as Figure 1.6 illustrates.

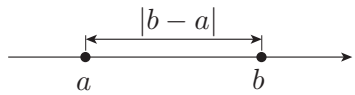


Figure 1.6 Real line showing $|b - a|$

Notice that $|b - a| = 0$ if, and only if, $a = b$.

Another important property of the modulus function is the *Triangle Inequality*.

Proposition 1.6 *Triangle Inequality in* $\mathbb{R}$

For all real numbers a and b,

$$|a + b| \leq |a| + |b|.$$

For example, with $a = -4$ and $b = 3$, $|-4 + 3| = |-1| = 1$ and $|-4| + |3| = 4 + 3 = 7$, so $|-4 + 3| \leq |-4| + |3|$.

Proof For all real numbers a and b, we have

$$\begin{aligned}|a+b|^2 &= (a+b)^2 \qquad \text{(by equation (3))}\\ &= a^2+2ab+b^2\\ &\le a^2+2|ab|+b^2 \qquad \text{(by equation (1))}\\ &= |a|^2+2\,|a|\,|b|+|b|^2 \qquad \text{(by equations (2) and (3))}\\ &= (|a|+|b|)^2.\end{aligned}$$

See Exercise 1.6.

So $|a+b|^2 \le (|a|+|b|)^2$. We now use the result that when p and q are non-negative numbers, $p^2 \le q^2$ implies $p \le q$. Hence

$$|a+b| \le |a|+|b|. \quad \blacksquare$$

Exercise 1.6

Prove that if p and q are non-negative numbers then

$$p^2 \le q^2 \implies p \le q.$$

Corollary 1.7

For all real numbers a, b and c,

$$|b-a| \le |b-c|+|c-a|.$$

Proof On noting that $b-a = (b-c)+(c-a)$, we apply Proposition 1.6 to find

$$|b-a| = |(b-c)+(c-a)| \le |b-c|+|c-a|. \quad \blacksquare$$

Exercise 1.7

Use the Triangle Inequality to show that:

(a) for all real numbers a, b and c, $|a+b+c| \le |a|+|b|+|c|$

(b) for all real numbers a and b, $|a-b| \le |a|+|b|$.

The following result provides a useful variant of the Triangle Inequality, known as the *backwards* or *reversed* form of the Triangle Inequality – or, more briefly, as the *Reverse Triangle Inequality*.

Proposition 1.8 *Reverse Triangle Inequality*

For all real numbers a and b,

$$|b-a| \ge \big||b|-|a|\big|.$$

Exercise 1.8

Prove Proposition 1.8.

1.2 Sequences

You will have met sequences of real numbers in your previous studies, but the idea of a sequence is such an important part of what follows that we now take a little time to remind you of the main ideas by recalling the notions of null sequences, convergent sequences, their combination rules and some useful theorems for determining whether a given sequence converges or diverges. We omit nearly all the proofs of these results; you will likely have seen them in an earlier module.

We begin by recalling the definition of a sequence of real numbers.

Definition 1.9 *Sequence of real numbers*

A **(real) sequence** is an unending ordered list of real numbers

$$a_1, a_2, a_3, \ldots.$$

The real number a_n ($n \in \mathbb{N}$) is the ***n*th term** of the sequence, and the whole sequence is denoted by (a_n).

A sequence (a_n) is **monotonic increasing** if for each $n \in \mathbb{N}$, $a_n \leq a_{n+1}$. It is **strictly monotonic increasing** if this inequality is strict for each n, $a_n < a_{n+1}$.

A sequence (a_n) is **monotonic decreasing** if for each $n \in \mathbb{N}$, $a_n \geq a_{n+1}$. It is **strictly monotonic decreasing** if this inequality is strict for each n, $a_n > a_{n+1}$.

Many texts use the notation $\{a_n\}$ instead of (a_n), but this makes it difficult to distinguish between the sequence (a_n) and its *set* of values, $\{a_n : n \in \mathbb{N}\}$.

Remarks

1. Alternatives to the notation (a_n) that we will sometimes use are $(a_n)_{n=1}^{\infty}$, $(a_n)_{n\in\mathbb{N}}$ (when we wish to emphasise the limits) and (in later chapters) $\mathbf{a}$.
2. By definition, unless we state otherwise, we assume that the first term of a sequence is a_1 (and not a_0).
3. The terms of a sequence (a_n) do not need to be different. For example, $-1, 1, -1, 1, \ldots$ and $1, 1, 1, 1, \ldots$ are sequences. A sequence for which all terms are the same is known as a **constant sequence**.
4. For any subset A of $\mathbb{R}$, we say that a sequence (a_n) is **in** A if $a_n \in A$ for each n. For example, the sequences $2, 4, 6, 8, \ldots$ and $1, 1, 1, 1, \ldots$ are in $\mathbb{N}$.
5. A sequence can also be viewed as a function and a more formal definition is then given by:

   a *(real) sequence* is a function $a\colon \mathbb{N} \to \mathbb{R}$ given by $n \mapsto a_n$.

We often define sequences by giving the first few terms and a formula for the nth term – for example,

$$2, 4, 6, \ldots, 2n, \ldots.$$

This sequence may also be written as

$$a_n = 2n \text{ for } n = 1, 2, 3, \ldots$$

or as $(2n)_{n=1}^{\infty}$ or simply (but possibly confusingly) as $(2n)$.

Sometimes the nth term is defined by a recursion formula, as with the Fibonacci sequence

$$1, 1, 2, 3, 5, 8, \ldots, \text{ where } a_{n+2} = a_{n+1} + a_n \text{ for } n \geq 1.$$

In this case, the first two terms of the sequence need to be specified in order to define the whole sequence.

Null sequences

If you look at the terms of the sequence

$$1, \tfrac{1}{2}, \tfrac{1}{3}, \ldots, \tfrac{1}{n}, \ldots$$

you will notice that as n gets large, the terms get closer to 0 (Figure 1.7). This sequence is an example from an important class of sequences, the class of *null sequences* – the sequences that *converge to* 0.

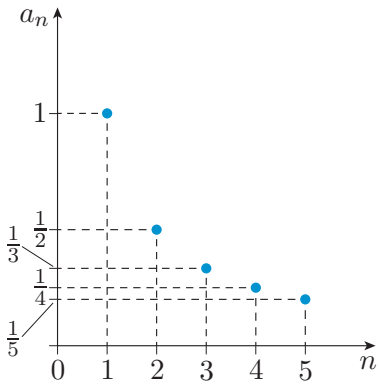


Figure 1.7 The first few terms of the sequence $\left(\frac{1}{n}\right)$

This notion of converging to 0 needs to be made precise. The idea is to show that for *any* positive number ε, we can find a natural number N such that all the terms of the sequence after a_N lie within the interval $(-\varepsilon, \varepsilon)$. A sequence is a null sequence if this property holds whatever positive number ε is taken, no matter how small.

For example, with the sequence $\left(\frac{1}{n}\right)$, suppose that $\varepsilon = 0.01$. If $\left(\frac{1}{n}\right)$ is a null sequence then there must be a value N for n for which all the subsequent terms in the sequence are within 0.01 of 0. In fact, this occurs when $N = 100$ because, from $n = 101$ onwards, $\frac{1}{n}$ is smaller than 0.01.

Often, we say simply that '(a_n) is null'.

Definition 1.10 *Null sequence*

A real sequence (a_n) is a **null sequence** if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n| < \varepsilon \quad \text{whenever } n > N.$$

We say that a null sequence **converges to 0**, and write

$$a_n \to 0 \text{ as } n \to \infty, \quad \text{or simply } a_n \to 0.$$

Remarks

1. In general, for a given sequence, the value of N depends on ε; usually, the smaller ε is, the larger N will be.

2. The definition uses the absolute value $|a_n|$ because terms of the sequence may be negative. As you will recall from Subsection 1.1, $|a_n| < \varepsilon$ means that $-\varepsilon < a_n < \varepsilon$: that is, on the graph of a_n against n, the terms of a null sequence eventually lie inside a band of width 2ε centred on 0, as Figure 1.8 illustrates.

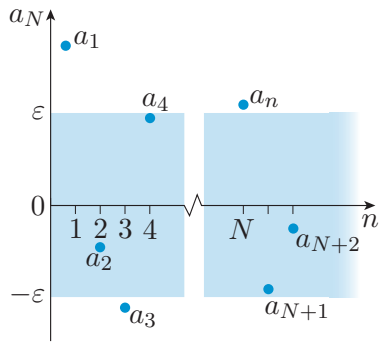


Figure 1.8 Null sequence

Worked Exercise 1.11

Prove that each of the following sequences is null:

(a) (a_n), where $a_n = 0$ for each n

(b) (b_n), where $b_n = \dfrac{1}{n^2}$ for each n

(c) (c_n), where $c_n = \dfrac{(-1)^n}{n^2}$ for each n.

Solution

(a) Let $\varepsilon > 0$ be given. Then $|a_n| = |0| = 0 < \varepsilon$ for each n.

Choose $N = 1$. Then $|a_n| < \varepsilon$ whenever $n > N$.

We deduce that (a_n) is null.

(b) Let $\varepsilon > 0$ be given. In this case $|b_n| = \left|\frac{1}{n^2}\right| = \frac{1}{n^2}$, and

$$\frac{1}{n^2} < \varepsilon \quad \text{if, and only if,} \quad \frac{1}{\sqrt{\varepsilon}} < n.$$

Choose $N = \text{int}\left(\frac{1}{\sqrt{\varepsilon}}\right)$, the largest integer less than or equal to $\frac{1}{\sqrt{\varepsilon}}$. Then, for each integer $n > N$, we have $n > \frac{1}{\sqrt{\varepsilon}}$, and hence

$$|b_n| = \frac{1}{n^2} < \varepsilon \quad \text{whenever } n > N.$$

We deduce that (b_n) is null.

(c) Let $\varepsilon > 0$ be given. Then

$$|c_n| = \left|\frac{(-1)^n}{n^2}\right| = \frac{1}{n^2} = b_n$$

for each n. Hence we choose N as in part (b), so that $|c_n| = |b_n| < \varepsilon$ whenever $n > N$.

We deduce that (c_n) is null.

Recall from Book C that $\text{int}(x)$ – also written $\lfloor x \rfloor$ – denotes the largest integer that does not exceed x. For example, $\text{int}(2) = 2$, $\text{int}(1.23) = 1$ and $\text{int}(-1.5) = -2$.

Exercise 1.9

Prove from the definition of a null sequence that the sequence (a_n) given by $a_n = \frac{1}{\sqrt{n}}$ is null.

To prove that a sequence is null, we have to show that whatever band of total width 2ε we make centred on 0, the terms of the sequence must eventually all lie inside that band. Thus to prove that a sequence (a_n) is *not* a null sequence, we need to show that this condition does not hold. This means that we need to find *just one* such band around 0 for which, no matter how far we progress in the sequence, there are always later terms of the sequence that lie outside that band.

More formally, we need to find one value of ε such that, no matter what choice of N we make, we can always find a natural number $n > N$ for which $|a_n| \geq \varepsilon$.

Worked Exercise 1.12

Prove that the sequence (a_n) given by $a_n = (-1)^n$ is not null.

Solution

Since the sequence is $-1, 1, -1, 1, \ldots$ we can take any $\varepsilon < 1$, say $\varepsilon = \frac{1}{2}$ (Figure 1.9). Then, since $|a_n| = 1$ for each n, $|a_n| \geq \varepsilon$ for each n. So there is no natural number N for which $|a_n| < \varepsilon$ for each $n > N$. Thus the sequence is not null.

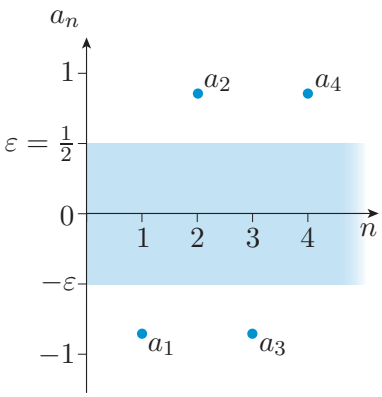


Figure 1.9 The sequence $a_n = (-1)^n$

Exercise 1.10

Prove that the constant sequence (a_n) given by $a_n = 2$ for each n is not null.

We do not prove that these sequences are null.

We record here some basic examples of null sequences.

Theorem 1.13 *Basic null sequences*

The following sequences are null:

- $\left(\frac{1}{n^p}\right)$, for any constant $p > 0$
- $(n^p c^n)$, for any constants $p \in \mathbb{R}$ and $|c| < 1$
- $\left(\frac{c^n}{n!}\right)$, for any constant $c \in \mathbb{R}$
- $\left(\frac{n^p}{n!}\right)$, for any constant $p \in \mathbb{R}$.

Examples:
$\left(\frac{1}{n^3}\right)$ when $p = 3$;
$\left(\frac{1}{2^n}\right)$ when $c = \frac{1}{2}$, $p = 0$;
$\left(\frac{2^n}{n!}\right)$ when $c = 2$;
$\left(\frac{n^{100}}{n!}\right)$ when $p = 100$.

This theorem is a consequence of the different growth rates of n^p, $\left(\frac{1}{c}\right)^n$ and $n!$ (for $|c| < 1$ and $p \in \mathbb{R}$): for n sufficiently large, it is always the case that

$$n^p \ll |c|^{-n} \ll n!.$$

It is often tricky to prove directly from the definition that a given sequence is null. However, there are several rules that allow us to combine null sequences to form new null sequences. They are summarised as follows.

Theorem 1.14 *Combination rules for null sequences*

Sum Rule	If (a_n) and (b_n) are null, then $(a_n + b_n)$ is null.
Multiple Rule	If (a_n) is null, then (λa_n) is null for each $\lambda \in \mathbb{R}$.
Product Rule	If (a_n) and (b_n) are null, then $(a_n b_n)$ is null.
Power Rule	If (a_n) is null, where $a_n \geq 0$ for each $n \in \mathbb{N}$, and $p > 0$, then (a_n^p) is null.

For example, since $\left(\frac{1}{n}\right)$ and $\left(\frac{1}{\sqrt{n}}\right)$ are (basic) null sequences, the Sum Rule implies that $\left(\frac{1}{n} + \frac{1}{\sqrt{n}}\right)$ is a null sequence, and the Product Rule (or the Power Rule) implies that $\left(\frac{1}{n^{3/2}}\right)$ is a null sequence.

Alternatively, we could note from our list of basic null sequences that $\left(\frac{1}{n^{3/2}}\right)$ is a sequence of the type $\left(\frac{1}{n^p}\right)$.

Another useful rule for proving that sequences are null is the *Squeeze Rule*. This rule uses the idea that, given a known null sequence, a sequence whose terms are eventually closer to 0 than the corresponding terms of the given sequence must also be null.

Lemma 1.15 *Squeeze Rule*

If (a_n) is null and if there is an $M \in \mathbb{N}$ such that $|b_n| \leq |a_n|$ for each $n > M$, then (b_n) is null.

Note that the Squeeze Rule can often be applied for each $n \in \mathbb{N}$ rather than for $n > M$, since if $|b_n| \leq |a_n|$ holds for each $n \in \mathbb{N}$ then it must hold for each $n > M$, whatever the value of $M \in \mathbb{N}$.

The sequence $\left(\frac{1}{2^{n^2}}\right)$ is null. This follows from the Squeeze Rule, since

$$\left|\frac{1}{2^{n^2}}\right| = \frac{1}{2^{n^2}} \leq \frac{1}{2^n} = \left|\frac{1}{2^n}\right| \quad \text{for each } n \in \mathbb{N},$$

and $\left(\frac{1}{2^n}\right)$ is a basic null sequence.

Worked Exercise 1.16

Let (a_n) be a null sequence and let $(b_n) = \left(\dfrac{a_n^2}{1+|a_n|}\right)$.
Use the combination and Squeeze Rules to prove that (b_n) is null.

Solution

We first observe that, by the Product Rule, (a_n^2) is null since it is the product of a_n with itself.

Now notice that, since $|a_n| \geq 0$ for each n,

$$\left|\frac{1}{1+|a_n|}\right| \leq 1$$

for each n, and so

$$\left|\frac{a_n^2}{1+|a_n|}\right| \leq |a_n^2|$$

for each n. Hence, by the Squeeze Rule, $(b_n) = \left(\dfrac{a_n^2}{1+|a_n|}\right)$ is null.

Exercise 1.11

Let (a_n) be a null sequence, let $\lambda \in \mathbb{R}$ with $\lambda \geq 0$ and let (b_n) be a sequence such that $|b_n| \leq \lambda |a_n|$ for each n. Use the combination and Squeeze Rules to prove that (b_n) is null.

If we were only able to investigate convergence of sequences tending to zero, then sequences would not be a very useful tool. Fortunately, we can use null sequences to define the convergence of sequences to values other than 0.

Definition 1.17 *Convergence and divergence*

A real sequence (a_n) **converges** to $l \in \mathbb{R}$ if the sequence $(a_n - l)$ is a null sequence: that is if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n - l| < \varepsilon \quad \text{whenever } n > N.$$

We say that l is the **limit** of the sequence (a_n) and write $a_n \to l$ as $n \to \infty$, or simply $a_n \to l$.

A sequence that does not converge to any real number is **divergent**.

Remarks

1. A sequence converges to l if, for each $\varepsilon > 0$, the terms of the sequence eventually lie inside a band of width 2ε centred on l, as Figure 1.10 illustrates.
2. If a sequence (a_n) does not converge to l, then we write $a_n \not\to l$ as $n \to \infty$. This means that either the sequence is divergent or it converges to some number other than l.
3. Convergent real sequences have unique limits. To show this, suppose that (a_n) is a sequence that converges to the real number a and to the real number b; then both $(a_n - a)$ and $(a_n - b)$ are null sequences. Moreover, the Triangle Inequality for the modulus function tells us that for each $n \in \mathbb{N}$,

$$|a - b| = |(a - a_n) - (b - a_n)| \leq |a - a_n| + |a_n - b|.$$

But both $|a - a_n| \to 0$ and $|a_n - b| \to 0$ as $n \to \infty$ and so by the Squeeze Rule, the constant sequence $(|a - b|)$ is a null sequence, implying $|a - b| = 0$. Thus $a = b$.

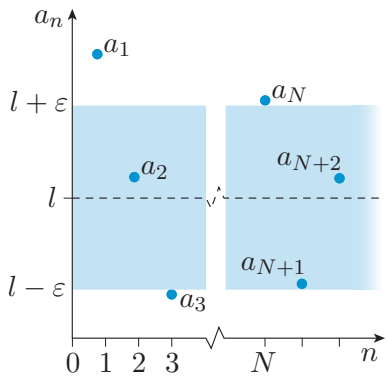


Figure 1.10 Sequence converging to l

Exercise 1.12

State conditions involving ε and N for a sequence (a_n) to be divergent, i.e. state conditions such that (a_n) does not converge to *any* limit l.

The result of Exercise 1.12 is summarised below.

(a_n) is a *divergent sequence* if for each $l \in \mathbb{R}$, there is an $\varepsilon > 0$ such that for each $N \in \mathbb{N}$, there is an $n > N$ with $|a_n - l| \geq \varepsilon$.

Worked Exercise 1.18

Prove that:

(a) the constant sequence $5, 5, 5, \ldots$ converges to 5

(b) the sequence $1, 2, 3, \ldots, n, \ldots$ is divergent.

Solution

(a) Let (a_n) be the constant sequence $5, 5, 5, \ldots$, and consider the sequence (b_n), where $b_n = a_n - 5$ for all n.

Then $b_n = 0$ for each n, and so is the constant sequence considered in Worked Exercise 1.11(a). Hence $(b_n) = (a_n - 5)$ is a null sequence, and so $a_n \to 5$ as $n \to \infty$.

(b) Here, $a_n = n$ for each n.

To show divergence, it is sufficient to find just *one* such $\varepsilon > 0$.

To prove divergence, we have to show that, given any $l \in \mathbb{R}$, there is an $\varepsilon > 0$ such that, no matter what choice of N we make, we can always find $n > N$ for which $|a_n - l| \geq \varepsilon$.

We take an arbitrary value of l and use the Reverse Triangle Inequality (Proposition 1.8) to find a lower bound on the size of $|a_n - l|$:

$$|a_n - l| = |n - l| \geq n - |l|. \tag{5}$$

For this particular sequence, we could choose any positive value we like for ε. We choose $\varepsilon = 1$ to make the algebra easier.

Let us take $\varepsilon = 1$. We will have demonstrated divergence if we can show that for any $N \in \mathbb{N}$, there is an $n > N$ for which $|a_n - l| \geq 1$, and equation (5) tells us that this will be true whenever

$$n - |l| \geq 1.$$

This inequality holds when $n \geq |l| + 1$. Thus, for each $n \geq |l| + 1$, we have

$$|a_n - l| \geq 1 = \varepsilon.$$

Here $\lceil l \rceil$ denotes the least integer greater than or equal to l.

Therefore, for each choice of N, if we take $n \in \mathbb{N}$ so that $n > \max\{N, |l| + 1\}$ (for example, take $n = \max\{N + 1, \lceil l \rceil + 1\}$, as shown in Figure 1.11), then $n > N$ and

$$|a_n - l| \geq 1 = \varepsilon.$$

We conclude that $a_n \not\to l$ as $n \to \infty$.

Since l is an arbitrary real number, (a_n) is a divergent sequence.

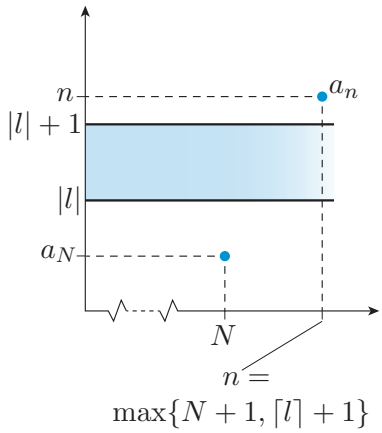


Figure 1.11 Demonstrating divergence

Remarks

1. Any constant sequence $l, l, l, \ldots$ converges to its constant value l.
2. Ideas similar to those used to show that $(n)_{n\in\mathbb{N}}$ is divergent can be used to show the following.

If (a_n) is a real null sequence, then $\left(\frac{1}{a_n}\right)$ is divergent.

Thus by taking the reciprocals of the terms of basic null sequences, we obtain divergent sequences.

As with null sequences, there are combination rules for combining convergent sequences to form new convergent sequences.

Theorem 1.19 *Combination rules for convergent sequences in* $\mathbb{R}$

If $a_n \to l$ and $b_n \to m$ as $n \to \infty$, then:

Sum Rule	$a_n + b_n \to l + m$ as $n \to \infty$
Multiple Rule	$\lambda a_n \to \lambda l$ as $n \to \infty$, for any $\lambda \in \mathbb{R}$
Product Rule	$a_n b_n \to lm$ as $n \to \infty$
Quotient Rule	$\frac{a_n}{b_n} \to \frac{l}{m}$ as $n \to \infty$, provided that $m \neq 0$
Squeeze Rule	if $l = m$ and $a_n \leq c_n \leq b_n$ for each n, then $c_n \to l$ as $n \to \infty$.

For the Quotient Rule, we are ignoring the fact that the ratio $\frac{a_n}{b_n}$ may not be defined for the first few terms; the condition that the limit of (b_n) is not zero guarantees that eventually b_n is not equal to zero.

These rules can be used, together with the basic null sequences, to prove that other sequences are convergent.

Worked Exercise 1.20

Prove that the following sequences converge, and find their limits.

(a) $(a_n) = \left((6 + 2^{-n})\left(2 - \frac{1}{n}\right)\right)$ (b) $(b_n) = \left(\frac{3n! + 2^n}{n^{50} + n!}\right)$

Solution

(a) Both (2^{-n}) and $(\frac{1}{n})$ are basic null sequences. Furthermore, by the Multiple Rule for null sequences (with $\lambda = -1$), the sequence $(-\frac{1}{n})$ is null. Also (6) and (2) are constant sequences and so converge to 6 and 2 respectively.

Hence, by the Sum Rule, $(6 + 2^{-n})$ converges to $6 + 0 = 6$ and $(2 - \frac{1}{n})$ converges to $2 + 0 = 2$.

Then, by the Product Rule, $(a_n) = ((6 + 2^{-n})(2 - \frac{1}{n}))$ converges to $6 \times 2 = 12$.

(b) Since $n!$ is the dominant term of b_n (that is, it grows faster than 2^n or n^{50} as n increases), we divide both the numerator and the denominator by $n!$. This gives

$$b_n = \frac{3 + 2^n/n!}{n^{50}/n! + 1}.$$

Both $\left(\frac{2^n}{n!}\right)$ and $\left(\frac{n^{50}}{n!}\right)$ are basic null sequences, and (3) and (1) are constant sequences. So, by the Sum Rule, the sequence $\left(3 + \frac{2^n}{n!}\right)$ converges to 3 and the sequence $\left(\frac{n^{50}}{n!} + 1\right)$ converges to 1. Hence, by the Quotient Rule, (b_n) converges to $\frac{3}{1} = 3$.

For sequences whose general term is expressed as a quotient, the technique of dividing both the numerator and the denominator by the *dominant term* is often useful when trying to prove convergence or divergence.

For constants c and p, $n!$ dominates c^n, which in turn dominates n^p.

Exercise 1.13

Prove that the following sequence converges, and find its limit.

$$\left(\frac{4n^2 + 1000^n}{n! + 7n^3}\right)$$

So far, in order to prove a given sequence is convergent, we have had to identify its limit first and then prove that the sequence converges to this limit. Sometimes this is difficult to do. For example, it is not obvious that the limit of the sequence (a_n) given by

$$a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

is $\frac{\pi^2}{6}$. So it is useful to have results that guarantee that a given sequence is convergent, even if we do not know its limit.

We now record a very useful result for concluding that a sequence is convergent, the *Monotone Convergence Theorem* for sequences, that does not require us to identify the limit. It is a powerful tool and we will return to the general problem of how to identify convergent sequences without finding their limits in a later chapter.

We expect that you will likely have met this result before and we do not prove it here.

See Book F, Chapter 23, *Completeness*.

To introduce it, we first remind you of the terms *upper bound* and *lower bound*.

Definition 1.21 *Upper and lower bounds*

Let A be a subset of real numbers. Then $b_l \in \mathbb{R}$ is a **lower bound** of A if for each $a \in A$, $b_l \leq a$.

Similarly, $b_u \in \mathbb{R}$ is an **upper bound** of A if for each $a \in A$, $a \leq b_u$.

Notice that there is no assumption in this definition that $b_l, b_u \in A$, nor does this terminology assume that these are the best possible choices of bound.

Theorem 1.22 *Monotone Convergence Theorem*

If (a_n) is a monotone increasing sequence that is bounded above by a real number b_u, then (a_n) is convergent with limit $l \leq b_u$.

If (a_n) is a monotone decreasing sequence that is bounded below by a real number b_l, then (a_n) is convergent with limit $l \geq b_l$.

Figure 1.12 illustrates these upper and lower bounds.

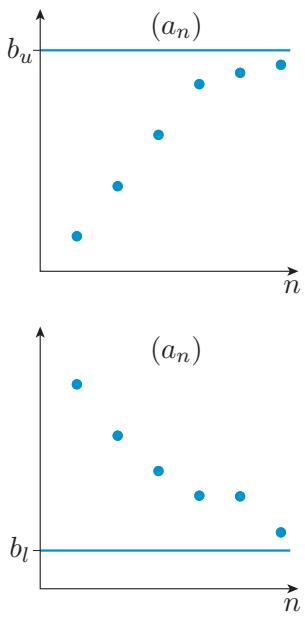


Figure 1.12 Upper and lower bounds for (a_n)

Worked Exercise 1.23

Prove that the sequence (a_n) given by

$$a_n = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2}$$

is convergent.

Solution

It is certainly the case that (a_n) is a monotonic increasing sequence, since for each $n \in \mathbb{N}$,

$$a_{n+1} - a_n = \frac{1}{(n+1)^2} > 0.$$

So if we can show that (a_n) is bounded above, then we can use the Monotone Convergence Theorem to conclude that (a_n) is a convergent sequence. But for each $n \geq 2$,

$$\frac{1}{n^2} \leq \frac{1}{n(n-1)} = \frac{1}{n-1} - \frac{1}{n}$$

and so for $n \geq 2$,

$$\begin{aligned} a_n &= 1 + \frac{1}{2^2} + \frac{1}{3^2} + \cdots + \frac{1}{n^2} \\ &\leq 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \\ &= 2 - \frac{1}{n}. \end{aligned}$$

$a_1 = 1 \leq 2.$

Hence (a_n) is a sequence that is bounded above by 2. We conclude by the Monotone Convergence Theorem that (a_n) is a convergent sequence with limit at most 2.

Next we recall a technique for investigating whether a sequence is divergent: looking at appropriate *subsequences*. The essential features of a subsequence are that all of its terms are in the original sequence, and that the terms occur in the same order as in the original sequence. In general, given a sequence (a_n), we can form many different subsequences.

Definition 1.24 *Subsequence*

The sequence $(a_{n_k})_{k=1}^{\infty}$ is a **subsequence** of the sequence (a_n) if $(n_k)_{k=1}^{\infty}$ is a strictly increasing sequence of positive integers, i.e. if $1 \leq n_1 < n_2 < n_3 < \cdots$.

In particular:

- when $n_k = 2k$, we have $(a_{2k})_{k=1}^{\infty}$, which is the **even subsequence** of (a_n)
- when $n_k = 2k - 1$, we have $(a_{2k-1})_{k=1}^{\infty}$, which is the **odd subsequence** of (a_n).

Remarks

1. Every sequence is a subsequence of itself, obtained by putting $n_k = k$, for $k = 1, 2, 3, \ldots$.
2. The even subsequence (a_{2k}) consists of the even-numbered terms $a_2, a_4, a_6, \ldots$ of the original sequence. For example, for the sequence $(a_n) = 1, 4, 9, 16, 25, 36, \ldots, n^2, \ldots$, the even subsequence (a_{2k}) is $4, 16, 36, \ldots, (2k)^2, \ldots$. Similarly, the odd subsequence consists of the odd-numbered terms $a_1, a_3, a_5, \ldots$.

Exercise 1.14

Show that (2^{k^2}) is a subsequence of (2^n).

The terms of (2^{k^2}) are $2, 2^4, 2^9, 2^{16}, \ldots$.

The following result relates the convergence of a given sequence to the convergence of its subsequences and provides us with a way to show that a sequence is divergent.

Proposition 1.25 *Convergence of subsequences*

If (a_n) is a convergent sequence with limit l, then every subsequence (a_{n_k}) is convergent with the same limit l.

Proposition 1.25 can be used to show that a given sequence is *divergent*, since it tells us the following.

1. If (a_n) has a divergent subsequence, then (a_n) is itself divergent.
2. If a sequence (a_n) has two convergent subsequences whose limits are different, then (a_n) is divergent.

Worked Exercise 1.26

Show that $(a_n) = ((-1)^n)$ is a divergent sequence.

Solution

We produce two subsequences of (a_n) that converge to different limits.

Consider the even subsequence $(a_{2k}) = ((-1)^{2k})$. Each term of this subsequence is 1, and so this subsequence converges to 1.

Now consider the odd subsequence $(a_{2k-1}) = ((-1)^{2k-1})$. Each term is -1, and so this subsequence converges to -1, a different limit. (See Figure 1.13.)

By Proposition 1.25, we conclude that $(a_n) = ((-1)^n)$ is divergent.

In Worked Exercise 1.12 we proved that this sequence is not a null sequence.

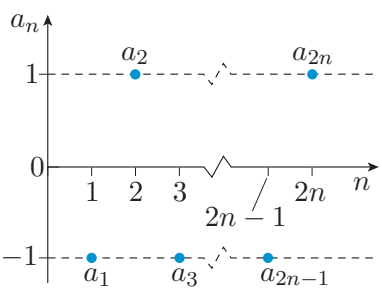


Figure 1.13 Subsequences of $(a_n) = ((-1)^n)$

Exercise 1.15

Determine whether the following sequence is convergent or divergent, justifying your answer.

$$\left((-1)^n \frac{1-n}{n}\right)$$

The final exercise in this subsection is an opportunity for you to practise testing sequences for convergence.

Exercise 1.16

Determine whether each of the following sequences converges. If so, what is its limit?

(a) $\left(\frac{3n^3}{n!} + \frac{7}{n^2}\right)$; (b) $\left((n+1)^2 - n^2\right)$; (c) $\left(\sqrt{n+1} - \sqrt{n}\right)$.

Hint: for part (c), you may find the following identity helpful:

$$a - b = (\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b}) \text{ for } a, b \geq 0.$$

2 Continuity on the real line

In this section, we recall the definition of a continuous function from $\mathbb{R}$ to $\mathbb{R}$. We use (convergent) sequences as our main tool for testing the continuity of functions.

2.1 Continuous and discontinuous functions

Consider the functions $f_1 : \mathbb{R} \to \mathbb{R}$ and $f_2 : \mathbb{R} \to \mathbb{R}$ illustrated in Figure 2.1 and given by

$$f_1(x) = x^2,$$

$$f_2(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

Intuitively, we would say that f_1 is continuous since its graph has no jumps or breaks in it, whereas f_2 is discontinuous since there is a jump at $x = 0$. We could go on to say that f_1 is continuous at every point $x \in \mathbb{R}$, whereas f_2 is discontinuous at $x = 0$.

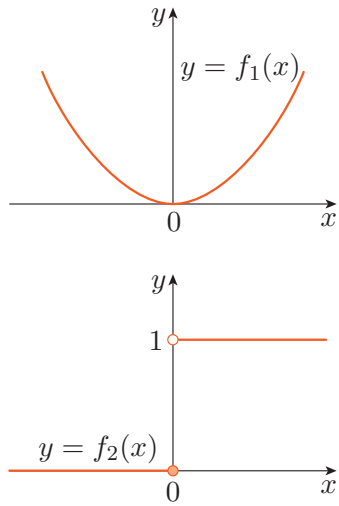


Figure 2.1 The functions $f_1(x)$ and $f_2(x)$

Our immediate objective is to capture this intuitive idea of a continuous function in a precise way. We do this by using the idea that if f is continuous at a point a and if x is a point near to a, then $f(x)$ should be near to $f(a)$. In other words, our definition should say in a precise way that

if x tends to a, then $f(x)$ tends to $f(a)$.

We make this notion exact by using the convergence of sequences.

Definition 2.1 *Continuity at a point*

Let $A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

Then f is **continuous** at $a \in A$ if:

whenever (x_n) is a sequence in A for which $x_n \to a$ as $n \to \infty$, then the sequence $(f(x_n))$ converges to $f(a)$.

If f does not satisfy this condition at $a \in A$ – that is, there is a sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$ but $f(x_n)$ does not converge to $f(a)$ – then we say that f is **discontinuous** at a.

Remarks

1. This definition of continuity is often called the **sequential definition of continuity**, or simply **sequential continuity**, to distinguish it from alternative definitions.
2. To show that f is continuous at a, we must show that $(f(x_n))$ converges to $f(a)$ for *every possible choice* of a sequence (x_n) that converges to a.
3. If the sequence $(f(x_n))$ converges to $f(a)$ for a *particular* sequence (x_n) that converges to a (in the domain A), then we often write

   $$f(x_n) \to f(a) \text{ as } x_n \to a.$$

   If the sequence $(f(x_n))$ converges to $f(a)$ for *every* sequence (x_n) that converges to a (in the domain A), then we often write

   $$f(x) \to f(a) \text{ as } x \to a.$$

Continuity as defined above is the property of a function *at a point*. It is easily extended to define continuity *on a set*.

Definition 2.2 *Continuity on a set*

Let $S \subseteq A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Remarks

1. If we simply say that a function $f: A \to \mathbb{R}$ is *continuous* (or *discontinuous*), by implication we mean that it is continuous (or discontinuous) on A.
2. To prove continuity on a set S, we need to show that the function f is continuous at *each* point a in S – if f is a 'hybrid' function made up of two or more functions, then this may require different arguments depending on the value of a.
3. To prove discontinuity on a set S, all that is needed is to prove it for a *single*, *specific* point in S.

We now show how the definition of continuity is used to prove that a function is continuous.

Worked Exercise 2.3

Prove that the function $f_1: \mathbb{R} \to \mathbb{R}$ given by

$$f_1(x) = x^2$$

is (a) continuous at 2 and (b) continuous on $\mathbb{R}$.

Solution

(a) The definition of continuity tells us that we must show that whenever (x_n) is a sequence in $\mathbb{R}$ that converges to 2, then the sequence $(f_1(x_n))$ converges to $f_1(2) = 4$.

Let (x_n) be a sequence (in $\mathbb{R}$) that converges to 2. We know that $f_1(x_n) = x_n^2$. Therefore, by the Product Rule for sequences,

$$f_1(x_n) = x_n \times x_n \to 2 \times 2 = 4 \text{ as } n \to \infty.$$

But $f_1(2) = 4$, and so $f_1(x_n) \to f_1(2)$ as $x_n \to 2$.

Hence f_1 is continuous at 2.

(b) This is a generalisation of the method in part (a).

Let a be a general point in $\mathbb{R}$. By the definition of continuity we must show that whenever (x_n) is a sequence in $\mathbb{R}$ that converges to a, then the sequence $(f_1(x_n))$ converges to $f_1(a) = a^2$.

Let (x_n) be a sequence (in $\mathbb{R}$) that converges to a. By the Product Rule for sequences,

$$f_1(x_n) = x_n^2 = x_n \times x_n \to a \times a = a^2 \text{ as } n \to \infty.$$

But $f_1(a) = a^2$, and so $f_1(x_n) \to f_1(a)$ as $x_n \to a$.

Hence f_1 is continuous at a.

Since a is an arbitrary point of $\mathbb{R}$, f_1 is continuous on $\mathbb{R}$.

Exercise 2.1

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = x^3 + 3$$

is continuous on $\mathbb{R}$.

To show that a function f is *discontinuous* at a point $a \in A$, we just need to find a single sequence (x_n) in A for which

$$x_n \to a \text{ but } f(x_n) \not\to f(a) \text{ as } n \to \infty.$$

Worked Exercise 2.4

Prove that the function $f_2\colon \mathbb{R} \to \mathbb{R}$ given by

$$f_2(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0, \end{cases}$$

is discontinuous at 0.

Solution

The definition of discontinuity tells us that we must find a sequence (x_n) in $\mathbb{R}$ for which $x_n \to 0$ as $n \to \infty$, but $f_2(x_n) \not\to f_2(0)$ as $n \to \infty$.

Since $f_2(0) = 0$ and $f_2(x) = 1$ for $x > 0$, we choose a sequence (x_n) that converges to 0 'from the right', so that $f_2(x_n) = 1$ for each n. One such sequence is given by $(x_n) = (\frac{1}{n})$, as shown in Figure 2.2. For this choice of sequence, we have

$$x_n = \tfrac{1}{n} \to 0 \text{ as } n \to \infty \text{ and yet } f_2(x_n) = 1 \text{ for each } n,$$

so $f_2(x_n) \to 1$ as $n \to \infty$. Since $f_2(0) = 0$ and $f_2(x_n) \not\to 0$,

$$f_2(x_n) \not\to f_2(0) \text{ as } n \to \infty,$$

and we conclude that f_2 is discontinuous at 0.

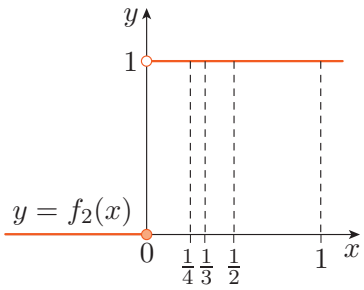


Figure 2.2 The sequence $(x_n) = (\frac{1}{n})$

Thus f_2 is discontinuous only at $a = 0$. This is what we would expect from looking at its graph, given in Figure 2.2.

Exercise 2.2

Prove that f_2 is continuous at a for each $a \neq 0$.

Worked Exercise 2.5

Let $f\colon \mathbb{R} \to \mathbb{R}$ be given by the rule

$$f(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q}, \\ 0, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Show that f is discontinuous at every $a \in \mathbb{R}$.

Solution

Let $a \in \mathbb{R}$.

We show that f is discontinuous at a by finding a sequence (x_n) such that

$$x_n \to a, \text{ but } f(x_n) \not\to f(a) \text{ as } n \to \infty.$$

If a is *rational*, then by definition $f(a) = 1$. We take a sequence of irrational numbers that converges to a – for example, let $x_n = a + \frac{1}{n}\sqrt{2}$. Then the combination rules for convergent sequences tell us that $x_n \to a + 0 \times \sqrt{2} = a$ as $n \to \infty$ but $f(x_n) = 0 \not\to f(a) = 1$, as $n \to \infty$.

We know that $a + \frac{1}{n}\sqrt{2}$ is irrational since if $a + \frac{1}{n} = \frac{p}{q}$ for some choice of integers p and q (with $q \neq 0$), then rearranging this expression would give $\sqrt{2} = n(\frac{p}{q} - a)$, a rational number.

Hence f is not continuous at a if $a \in \mathbb{Q}$.

If a is *irrational*, then $f(a) = 0$, and we choose a sequence of *rational* numbers (x_n) that converges to a. (Such a sequence exists because the density property of the reals tells us that between any two distinct real numbers there is a rational number; in particular, we can find a rational number x_n in the interval $(a, a + \frac{1}{n})$.)

The Squeeze Rule then guarantees that the sequence of rationals (x_n) converges to a, since $a + \frac{1}{n} \to a$ and $a \to a$ as $n \to \infty$.

Thus $x_n \to a, \text{ but } f(x_n) = 1 \not\to f(a) = 0, \text{ as } n \to \infty.$

Hence f is not continuous at a if $a \notin \mathbb{Q}$.

We deduce that f is discontinuous at every point in $\mathbb{R}$.

Let us now see what our definition tells us about the continuity of a function whose behaviour is less obvious.

Worked Exercise 2.6

Let $f\colon \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} \cos(\pi/x), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

Decide whether f is continuous at 0, and justify your decision.

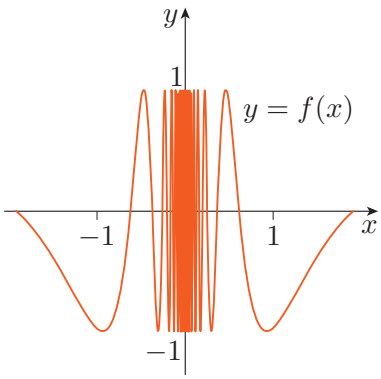


Figure 2.3 The graph of f for Worked Exercise 2.6

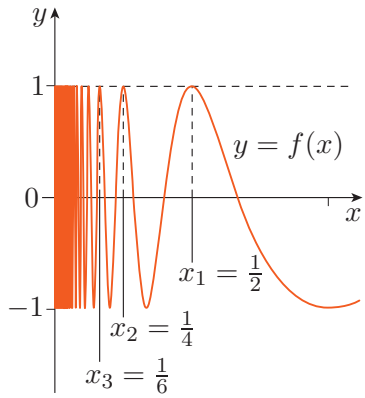


Figure 2.4 Choosing the sequence (x_n)

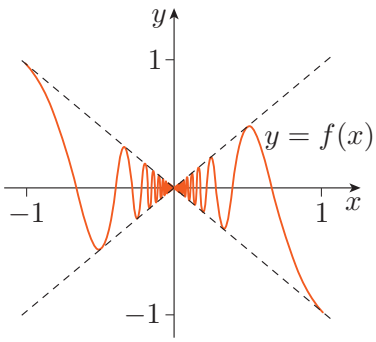


Figure 2.5 The graph of f for Exercise 2.3

Solution

As Figure 2.3 shows, the graph of f oscillates very rapidly close to the origin, and there are points x, arbitrarily close to the origin, where $\cos(\pi/x)$ seems to take values that are far from $0 = f(0)$. We expect that f is *discontinuous* at 0.

In order to prove this, we need to find a single sequence (x_n) in $\mathbb{R}$ for which

$$x_n \to 0 \text{ but } f(x_n) \not\to f(0) = 0.$$

We choose a sequence (x_n) of points for which $\cos(\pi/x_n) = 1$ (that is, the maximum points of the oscillation), so that $f(x_n) = \cos(\pi/x_n) = 1$ for each n and hence $f(x_n) \to 1 \neq 0$.

Now $\cos a = 1$ whenever a is of the form $2\pi n$, where n is an integer. It follows that $\cos(\pi/x) = 1$ whenever $\pi/x = 2\pi n$, i.e. whenever $x = \frac{1}{2n}$. So if we take (x_n) to be the sequence given by $x_n = \frac{1}{2n}$, as shown in Figure 2.4, then

$$f(x_n) = \cos\left(\frac{\pi}{x_n}\right) = \cos\left(\frac{\pi}{1/(2n)}\right) = \cos 2\pi n = 1 \quad \text{for each } n.$$

Moreover, $x_n \to 0$ as $n \to \infty$.

So we have found a sequence (x_n) for which $x_n \to 0$ but $f(x_n) \not\to f(0)$.

Thus f is discontinuous at 0.

Exercise 2.3

Prove that the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x\cos(\pi/x), & \text{if } x \neq 0, \\ 0, & \text{if } x = 0, \end{cases}$$

is continuous at 0. (This function is shown in Figure 2.5.)

Exercise 2.4

Let $f\colon \mathbb{R} - \{1\} \to \mathbb{R}$ be given by $f(x) = \dfrac{x^2 - 1}{x - 1}$.

(a) Suppose that (x_n) is a sequence in $\mathbb{R} - \{1\}$ with $x_n \to 1$ as $n \to \infty$. Show that $(f(x_n))$ is a convergent sequence and find its limit.

(b) Deduce that if $g\colon \mathbb{R} \to \mathbb{R}$ is given by

$$g(x) = \begin{cases} f(x), & \text{if } x \neq 1, \\ 2, & \text{if } x = 1, \end{cases}$$

then g is continuous at $a = 1$.

The discontinuity at $a = 1$ for the function f in Exercise 2.4 is called a *removable* discontinuity because, by making an appropriate change to the definition of f (as in Exercise 2.4(b)), it is possible to define a continuous function that is identical to f everywhere except at the point $a = 1$. This contrasts with the situation for the function f of Worked Exercise 2.6, where the point $a = 0$ is known as an *essential* discontinuity – it is not possible to change the definition of this function at $a = 0$ in such a way as to make it continuous there. (We do not prove this, but the argument is similar to that given in Worked Exercise 2.6.)

2.2 Rules for continuous functions

It would be very laborious if every time we wished to show a given function was continuous, we had to go back to 'first principles' and work directly from the definition of continuity. Fortunately, by using a list of standard continuous functions together with combination rules for continuous functions (that are similar to those we have for convergent sequences), we can frequently avoid recourse to direct use of the definition of continuity.

We start by recording the combination rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$. They follow directly from the combination rules for convergent sequences in $\mathbb{R}$.

This should not be a surprise: continuity is defined in terms of convergent sequences, so one would expect many results about convergent sequences to have analogues for continuous functions.

Combination rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}$ and suppose $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ are continuous at $a \in A$.

Then the following functions are also continuous at a:

Sum Rule	$f + g\colon A \to \mathbb{R}$, defined by $(f + g)(x) = f(x) + g(x)$
Multiple Rule	$\lambda f\colon A \to \mathbb{R}$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$
Product Rule	$fg\colon A \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$
Quotient Rule	$f/g\colon A - \{x : g(x) = 0\} \to \mathbb{R}$, defined by $(f/g)(x) = f(x)/g(x)$, provided $g(a) \neq 0$.

Another useful result tells us that restricting the domain of a continuous function to a smaller set still gives us a continuous function. Before we state this result, we need a definition.

Definition 2.7 *Restricted function*

If the domain of a function $f\colon A \to B$ is restricted to a set C, where $C \subseteq A$, then the resulting **restricted function**, known as the **restriction of f to C**, is denoted by $f|_C\colon C \to B$, and given by $f|_C(x) = f(x)$ for each $x \in C$.

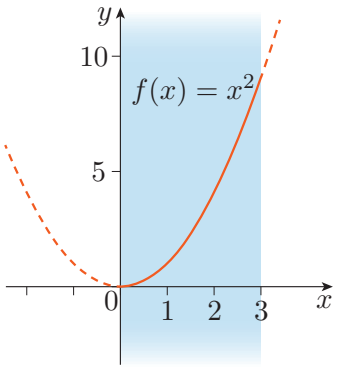


Figure 2.6 The function $f(x) = x^2$ restricted to $[0, 3]$

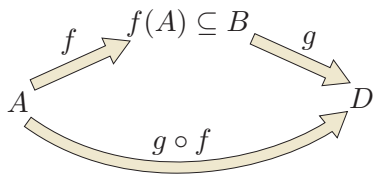


Figure 2.7 Composition

Here $A = B = C = \mathbb{R}$, $D = [-1, 1]$ and $f(A) = \mathbb{R} \subseteq B$.

Proposition 2.8 *Restriction Rule for continuous functions on $\mathbb{R}$*

Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$ and let $B \subseteq A$. If f is continuous at $b \in B$, then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous at b.

In particular, if f is continuous on A, then $f|_B$ is continuous on B.

For example, the function $f(x) = x^2$ is continuous on $\mathbb{R}$. Let $B = [0, 3]$. Then $f|_B\colon B \to \mathbb{R}$ is f restricted to the interval $[0, 3]$ and is continuous on $[0, 3]$, as Figure 2.6 illustrates.

Our final useful result about combining continuous functions allows us to compose continuous functions to form another continuous function. Before stating this result, we remind you of the definition of the *composition* of two functions.

Definition 2.9 *Composition*

If $f\colon A \to C$ and $g\colon B \to D$ are functions with $f(A) \subseteq B$, then the **composed function** or **composition** $g \circ f\colon A \to D$ is given by $(g \circ f)(x) = g(f(x))$.

Note that the condition $f(A) \subseteq B$ is needed so that the image set of the function f lies inside the domain of g (Figure 2.7). Some definitions of composition omit this condition, but we will require it to hold.

For example, let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^3$ and let $g\colon \mathbb{R} \to [-1, 1]$ be given by $g(x) = \sin x$. Then $g \circ f\colon \mathbb{R} \to [-1, 1]$ is given by $(g \circ f)(x) = \sin(x^3)$.

Proposition 2.10 *Composition Rule*

Let A, $B \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be continuous on A and B respectively, with $f(A) \subseteq B$. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

We prove a more general version of this result in Chapter 14 (Proposition 4.4) and do not prove it here.

These rules enable us to use a collection of basic continuous functions to form new functions that must then also be continuous on their domains.

Basic continuous functions

The following functions are continuous:

- polynomials and rational functions
- $f(x) = |x|$
- $f(x) = \sqrt{x}$
- the trigonometric functions (sine, cosine and tangent) and their inverses (restricted to appropriate domains)
- the logarithmic function
- the exponential function.

Note that by convention, the domains of these basic continuous functions are taken to be the largest subset of the reals for which the functions are defined. For example, the domain of a rational function excludes only those points where the denominator is zero, the domain of $f(x) = \sqrt{x}$ is $[0, \infty)$, and the domain of tan excludes the points $(2n + 1)\pi/2$ for all $n \in \mathbb{Z}$.

The application of the combination rules to functions constructed from basic continuous functions is illustrated in the following worked exercise.

Worked Exercise 2.11

Using the above rules for continuous functions, show that the function $f\colon [-\pi, \pi] \to \mathbb{R}$ given by

$$f(x) = e^x + 3\cos 2x$$

is continuous on $[-\pi, \pi]$.

Solution

The functions $x \mapsto 2x$, $x \mapsto e^x$ and $x \mapsto \cos x$ are all basic continuous functions, defined on the whole of $\mathbb{R}$.

By the Restriction Rule, both $x \mapsto e^x$ and $x \mapsto 2x$ are continuous functions on the interval $[-\pi, \pi]$.

The image set of $x \mapsto 2x$ restricted to $[-\pi, \pi]$ is $[-2\pi, 2\pi] \subseteq \mathbb{R}$. Thus, by the Composition Rule, $x \mapsto \cos 2x$ is continuous on $[-\pi, \pi]$. So, by the Multiple Rule, $x \mapsto 3\cos 2x$ is continuous on $[-\pi, \pi]$.

Finally, by the Sum Rule, f is continuous on $[-\pi, \pi]$.

Exercise 2.5

Show that $f\colon (0, \infty) \to \mathbb{R}$ given by

$$f(x) = 2\sin(1/x) - 3e^{-4x^3}$$

is continuous on $(0, \infty)$.

Exercise 2.6

Let $f: \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that the function $|f|: \mathbb{R} \to \mathbb{R}$ given by $|f|(x) = |f(x)|$ is continuous on $\mathbb{R}$.

2.3 Continuity theorems

In this subsection we remind you of three important theorems concerning continuous functions from $\mathbb{R}$ to $\mathbb{R}$. We do not prove these theorems here because we expect that you will have encountered them in your earlier studies. These theorems give us important information about the properties of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. One of our goals when we extend the notion of continuity to more general settings will be to see how far these theorems still hold true.

The proofs of the more general results we obtain in later chapters will also give proofs of the results we discuss here.

All three theorems concern continuous functions whose domain is a **closed bounded interval**, that is, an interval of the form

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$$

for real numbers a and b.

The Intermediate Value Theorem

Our first theorem says that continuous functions do not 'omit any values', and so there cannot be any jumps or gaps in their graphs. Thus, when f is continuous on a closed interval and attains two particular values, it must also take all the values between them. For example, let f be continuous on $[a, b]$, and suppose that $f(a) < 0$ and $f(b) > 0$. Then the theorem asserts that there must be at least one point $c \in (a, b)$ for which $f(c) = 0$ – that is, the graph of f must cross the x-axis, as Figure 2.8 illustrates.

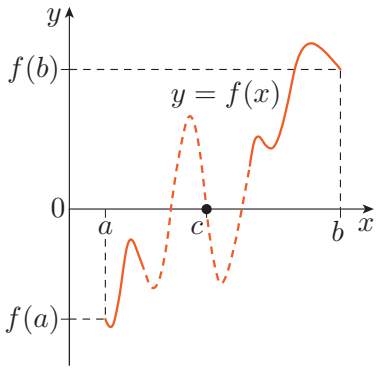


Figure 2.8 Graph of continuous function on $[a, b]$

Theorem 2.12 *Intermediate Value Theorem*

Let $f: [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ with $f(a) \neq f(b)$ and let k be any number lying between $f(a)$ and $f(b)$. Then there exists a number $c \in (a, b)$ such that $f(c) = k$.

Remarks

1. Notice that the theorem says nothing about the case when $f(a) = f(b)$.
2. By making use of the Restriction Rule (Proposition 2.8), this theorem can be adapted to apply to all intervals and not just to ones that are closed and bounded. To see this, suppose g is a continuous function from an interval I to $\mathbb{R}$, and a and b lie in I (with $a < b$). Then, by the Restriction Rule, $f: [a, b] \to \mathbb{R}$ given by $f(x) = g(x)$ for $x \in [a, b]$ is a continuous function on $[a, b]$ and so satisfies the hypotheses of the

I is of one of the following forms: $\mathbb{R}$, (r, s), $[r, s)$, $(r, s]$, $[r, s]$, $(-\infty, s)$, $(-\infty, s]$, (r, ∞), $[r, \infty)$ for $r, s \in \mathbb{R}$.

Intermediate Value Theorem. We conclude that if $g\colon I \to \mathbb{R}$ is continuous on the interval I and $a, b \in I$ are given, then whenever k lies between $g(a)$ and $g(b)$, there is $c \in (a, b)$ such that $g(c) = k$.

3. The Intermediate Value Theorem does not hold if the domain of f is not an interval. For example, consider the function $f(x) = \frac{1}{x}$ restricted to the set $[-1, 1] - \{0\}$, as shown in Figure 2.9. This function is continuous on $[-1, 1] - \{0\}$ and has $f(-1) < 0$ and $f(1) > 0$, but there is no value of $x \in [-1, 1] - \{0\}$ for which $f(x) = 0$.

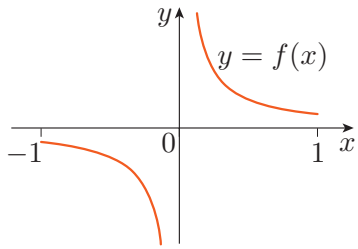


Figure 2.9 Graph of $f(x) = \frac{1}{x}$ on $[-1, 1] - \{0\}$

Worked Exercise 2.13

Let $f\colon [-\pi, \pi] \to \mathbb{R}$ be the function given by

$$f(x) = e^x + 3\cos 2x.$$

Prove that there is at least one point $c \in [-\pi, \pi]$ for which $f(c) = 0$.

Solution

We saw in Worked Exercise 2.11 that f is continuous on $[-\pi, \pi]$.

To show that there is a point $c \in [-\pi, \pi]$ for which $f(c) = 0$, using the Intermediate Value Theorem, we need to find an interval $[a, b]$ for which $f(a) < 0$ and $f(b) > 0$. We cannot use the endpoints of $[-\pi, \pi]$ because both $f(-\pi)$ and $f(\pi)$ are positive (see Figure 2.10). So we need to restrict our attention to a suitable subinterval of $[-\pi, \pi]$. The set $[-\frac{\pi}{2}, 0]$ is a closed interval contained in $[-\pi, \pi]$, and the Restriction Rule implies that f is continuous on it. Moreover,

$$f(-\tfrac{\pi}{2}) = e^{-\frac{\pi}{2}} + 3\cos(-\pi) = e^{-\frac{\pi}{2}} - 3 < 0,$$

$$f(0) = e^0 + 3\cos 0 = 1 + 3 = 4 > 0.$$

Since the conditions of the Intermediate Value Theorem are satisfied by f on $[-\frac{\pi}{2}, 0]$, we deduce that there is a number c (with $-\frac{\pi}{2} < c < 0$) for which $f(c) = 0$.

We could also have used the interval $[-\pi, -\frac{\pi}{2}]$, as can be seen from Figure 2.10.

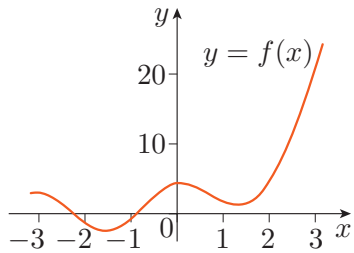


Figure 2.10 Graph of $f(x) = e^x + 3\cos 2x$ on $[-\pi, \pi]$

$e^{-\frac{\pi}{2}} < e^0 = 1.$

The Intermediate Value Theorem is an *existence result*: it tells us that there is a point in the interval whose image under f is 0, but it does not tell us how to find such a point. Thus, although we can deduce that the function $f\colon [-\pi, \pi] \to \mathbb{R}$ given by $f(x) = e^x + 3\cos 2x$ takes the value 0 at some point(s) in the interval $[-\frac{\pi}{2}, 0]$, we must use different methods to find the value(s) of x for which this is true.

The unique solution to $e^x + 3\cos 2x = 0$ in $[-\frac{\pi}{2}, 0]$ is approximately $x = -0.8564$.

Although the theorem does not provide a method for finding a point c at which a function f takes an intermediate value k, it is nevertheless very useful. It enables us to know in advance whether such a point c exists, and it applies to *all* continuous functions defined on a closed bounded interval.

Exercise 2.7 *A fixed-point theorem*

Let $f\colon [0,1] \to [0,1]$ be a continuous function. Use the Intermediate Value Theorem to show that there is a number $k \in [0,1]$ such that $f(k) = k$.

Hint: consider the function $g\colon [0,1] \to \mathbb{R}$ defined by $g(x) = f(x) - x$.

Exercise 2.7 shows that each continuous function $f\colon [0,1] \to [0,1]$ has a *fixed point*, a point that maps to itself under f.

The Boundedness and Extreme Value Theorems

The next two theorems involve not only closed bounded intervals but *boundedness* on $\mathbb{R}$ more generally. So, before we discuss the theorems, we remind you of the definition of boundedness on $\mathbb{R}$.

Definition 2.14 *Boundedness*

A set $A \subset \mathbb{R}$ is **bounded** if there exists a real number M such that $|x| \leq M$ for each $x \in A$.

A function $f\colon A \to \mathbb{R}$ is **bounded** on A if there is a real number M such that $|f(x)| \leq M$ for each $x \in A$.

Notice that closed bounded intervals $[a,b]$ are bounded sets.

Open intervals of the form (a,b) for real numbers a and b are also bounded sets.

Our first theorem is an immediate consequence of the second, but is easier to state and so we give it first.

Theorem 2.15 *Boundedness Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$. Then f is bounded on $[a,b]$. That is, there is a real number M such that $|f(x)| \leq M$ for each $x \in [a,b]$.

This theorem states that continuous functions on closed bounded intervals must be bounded; that is, the image set of the closed bounded interval $[a,b]$ under f is a bounded set. Note that this is not necessarily the case if the domain is not bounded: for example, the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x$ is a continuous function, and the image set of $\mathbb{R}$ under f is again $\mathbb{R}$, an *un*bounded set.

Similarly, if f is a continuous function that has a domain which is an open interval, then f need not be bounded: consider $f\colon (0,1) \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$.

Our next result tells us that a continuous function on a closed bounded interval is not only bounded but also attains its maximum and minimum values.

Theorem 2.16 *Extreme Value Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$. Then there are numbers $c, d \in [a,b]$ such that

$$f(c) \leq f(x) \leq f(d) \quad \text{for each } x \in [a,b].$$

Remarks

1. The theorem does not hold if either of the hypotheses (i.e. that f is continuous, and that the interval is closed and bounded) is dropped.
2. The number c is such that $f(c) \leq f(x)$ for all $x \in [a,b]$ – that is, $f(c)$ is the *minimum* value of the function in the interval. Similarly, $f(d)$ is the *maximum* value of the function in the interval.

Example 2.17

Consider $f\colon [-\pi,\pi] \to \mathbb{R}$ defined by $f(x) = e^x + 3\cos 2x$. Because f is continuous on $[-\pi,\pi]$ (see Worked Exercise 2.11), we can use the Extreme Value Theorem to deduce that there are points c and d in $[-\pi,\pi]$ such that $f(c) \leq f(x) \leq f(d)$ for all $x \in [-\pi,\pi]$. As Figure 2.11 indicates, c lies near -1.6 whilst $d = \pi$.

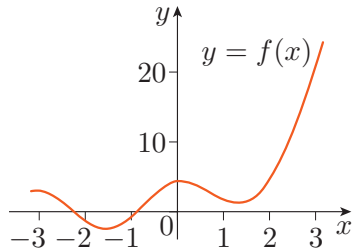


Figure 2.11 Graph of $f(x) = e^x + 3\cos 2x$ on $[-\pi,\pi]$

The Extreme Value Theorem is another example of an existence result: again, it does not tell us how to find the points whose existence it guarantees – we have to use other methods to do this.

Note that the Intermediate Value, Boundedness and Extreme Value Theorems hold for *all continuous functions* on *all closed bounded intervals*. We conclude that they are the result of an interaction between continuity and closed bounded intervals. When we look later at the properties of more general spaces, a generalisation of the idea of closed bounded intervals will give rise to the notion of 'compactness' – see Book F, Chapter 22.

Exercise 2.8

Let the function $f\colon [-1,1] - \{0\} \to \mathbb{R}$ be given by $f(x) = \frac{1}{x}$.

Explain why the Boundedness Theorem and the Extreme Value Theorem do not apply to this function.

3 Continuity on the plane

One of the main ways in which mathematics develops is by taking ideas from one situation and extending or generalising them into new ones. This process involves isolating some key aspects and ignoring others. We end this chapter by seeing whether we can use the definition of continuity for functions on the real line to come up with a definition of continuity for real-valued functions defined on the plane: functions with domain the plane, $\mathbb{R}^2$, and codomain $\mathbb{R}$. This will involve extending the ideas of distance, convergent sequences and continuity, which we developed in Sections 1 and 2, to make them appropriate for functions from $\mathbb{R}^2$ to $\mathbb{R}$. In the next chapter we will continue this process by first extending these ideas to spaces of higher dimensions and then considering even more general ideas of distance.

3.1 Functions from $\mathbb{R}^2$ to $\mathbb{R}$

In $\mathbb{R}^2$, we work with the familiar coordinate geometry of two dimensions. However, instead of using the common notation (x, y) for points in the plane, we use the notation (x_1, x_2) since this subscript notation is easy to extend to spaces of higher dimensions. We also use boldface letters such as $\mathbf{x}$ for the point (x_1, x_2), and $\mathbf{0}$ for the point $(0, 0)$. We refer to the point $\mathbf{0}$ as the **origin** of $\mathbb{R}^2$. For any two points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$, we define

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2) \quad \text{and} \quad \mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2).$$

In specifying the rule of a function f with domain $\mathbb{R}^2$ we use the notation $f(\mathbf{x})$ or $f(x_1, x_2)$. Two examples of functions from $\mathbb{R}^2$ to $\mathbb{R}$ are given by

$$f(x_1, x_2) = 20 - (x_1^2 + x_2^2) \quad \text{and} \quad g(x_1, x_2) = \frac{x_1^2 x_2}{x_1^2 + x_2^2 + 1}.$$

Although, strictly speaking, $f(\mathbf{x}) = f((x_1, x_2))$, we omit one pair of brackets and write $f(x_1, x_2)$.

Just as we represent (the graphs of) functions from $\mathbb{R}$ to $\mathbb{R}$ by curves in the plane, the graphs of functions from $\mathbb{R}^2$ to $\mathbb{R}$ are represented by *surfaces* in three-dimensional space. Above each point (x_1, x_2) in the domain $\mathbb{R}^2$, which we draw as a horizontal plane, we mark a point whose third coordinate x_3 is the function value, $f(x_1, x_2)$. Thus, using the examples given, above the point $(3, 2)$ we mark the point $f(3, 2) = 20 - 13 = 7$. As (x_1, x_2) ranges over different values in the domain $\mathbb{R}^2$, the values of $x_3 = f(x_1, x_2)$ define a surface in space. The graphs of the functions f and g are shown as surfaces in three-dimensional space in Figure 3.1.

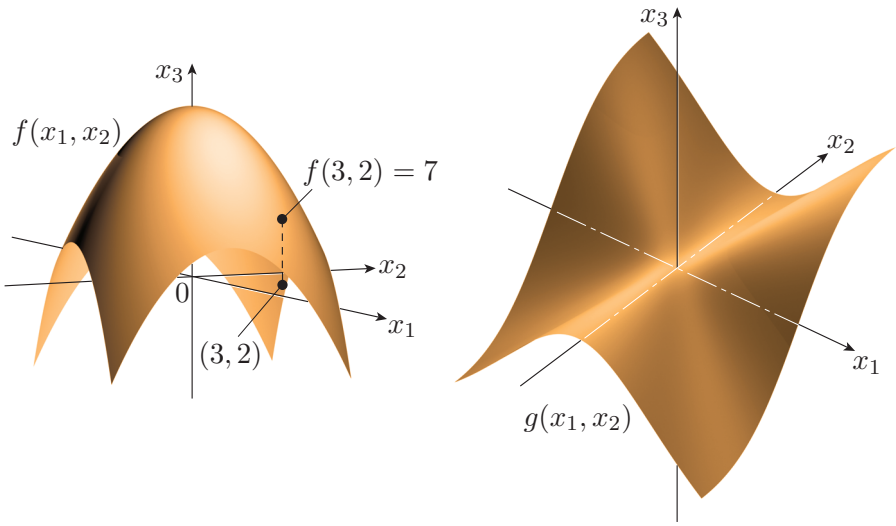


Figure 3.1 Graphs of $f, g\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x_1, x_2) = 20 - (x_1^2 + x_2^2)$ and $g(x_1, x_2) = \dfrac{x_1^2 x_2}{x_1^2 + x_2^2 + 1}$

As a first step in considering the continuity of functions on $\mathbb{R}^2$, we follow the development in Sections 1 and 2 and look at sequences in $\mathbb{R}^2$.

3.2 Sequences in $\mathbb{R}^2$

In Section 2 we defined continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ in terms of the convergence of appropriate sequences of real numbers. We now discuss sequences of points in the plane $\mathbb{R}^2$, and show how to define convergence for these. This will allow us to define the continuity of functions from $\mathbb{R}^2$ to $\mathbb{R}$ (in the next subsection).

Definition 3.1 *Sequence in* $\mathbb{R}^2$

A **sequence in** $\mathbb{R}^2$ is an unending ordered list of points in $\mathbb{R}^2$:

$$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots.$$

The point $\mathbf{a}_n = (a_{1,n}, a_{2,n})$ is the ***n*th term** of the sequence, and the whole sequence is denoted by $(\mathbf{a}_n)$, $(\mathbf{a}_n)_{n=1}^{\infty}$ or $(\mathbf{a}_n)_{n\in\mathbb{N}}$.

Remarks

1. Each term in the sequence is a point in the two-dimensional plane $\mathbb{R}^2$. This explains the double subscript: the point $\mathbf{a}_n$ has first coordinate $a_{1,n}$ and second coordinate $a_{2,n}$. So the nth term $\mathbf{a}_n$ is in $\mathbb{R}^2$, whereas its coordinates $a_{1,n}$ and $a_{2,n}$ are real numbers.
2. We could also write the sequence $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots$ as

$$(a_{1,1}, a_{2,1}),\ (a_{1,2}, a_{2,2}),\ (a_{1,3}, a_{2,3}),\ \ldots.$$

Example 3.2

Examples of sequences in the plane include the sequence

$$(\mathbf{a}_n) \text{ where } \mathbf{a}_n = (\cos n, \sin n),$$

consisting of certain points on the unit circle, and the sequence

$$(\mathbf{b}_n) \text{ where } \mathbf{b}_n = \left(\tfrac{1}{n}, n\right),$$

consisting of certain points on the graph of $f\colon (0,\infty) \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$ (see Figure 3.2).

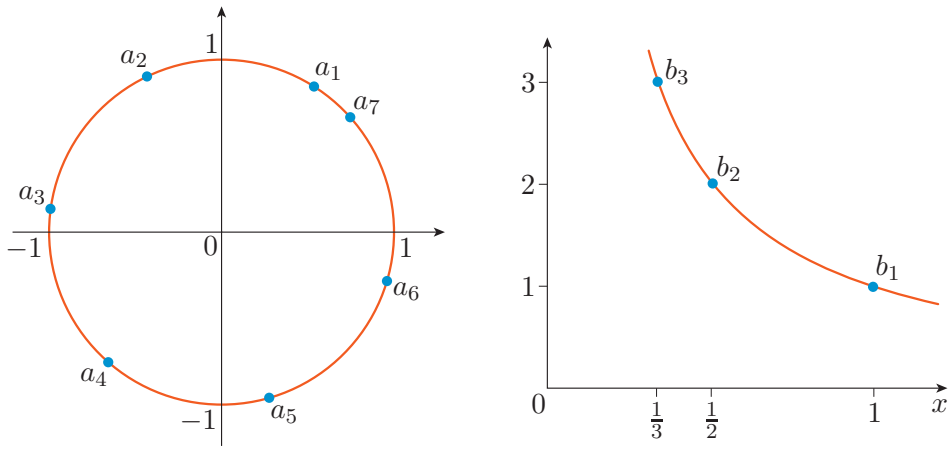


Figure 3.2 Examples of sequences in the plane

For sequences in $\mathbb{R}$, we have the notion of convergence: recall that a sequence (a_n) of real numbers is convergent (with limit l) if for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $|a_n - l| < \varepsilon$ whenever $n > N$. Notice that $|a_n - l|$ measures the distance between the two real numbers a_n and l, and that if we replaced this term with the distance between pairs of (appropriate) points in the plane, we would (potentially) have a definition of convergence for sequences of points in the plane. This motivates the following brief digression.

Distance in the plane

We saw in Section 1 that on the real line $\mathbb{R}$, the distance between two points a and b is given by $|b - a|$. In $\mathbb{R}^2$ we want the distance between two points to be defined in the usual way: for example, as Figure 3.3 illustrates, the distance between the points $(3, -1)$ and $(7, 2)$ is

$$\sqrt{(7-3)^2 + (2-(-1))^2} = \sqrt{4^2 + 3^2} = 5.$$

This distance, which is calculated by Pythagoras's Theorem, is called the *Euclidean distance* in $\mathbb{R}^2$.

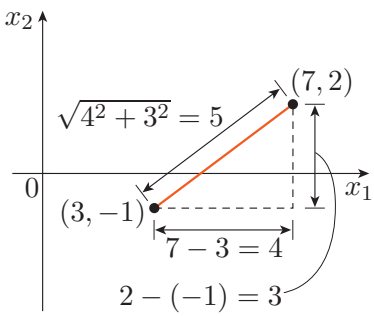


Figure 3.3 Distance between $(3, -1)$ and $(7, 2)$

Definition 3.3 *Euclidean distance in* $\mathbb{R}^2$

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$ is given by the formula

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Putting $\mathbf{b} = \mathbf{0} = (0, 0)$ in the definition of $d^{(2)}(\mathbf{a}, \mathbf{b})$ gives

$$d^{(2)}(\mathbf{a}, \mathbf{0}) = \sqrt{a_1^2 + a_2^2},$$

which is the distance between $\mathbf{a}$ and the origin, $\mathbf{0}$.

If you think of $\mathbf{a}$ as a vector, then $d^{(2)}(\mathbf{a}, \mathbf{0})$ is the *magnitude* of $\mathbf{a}$, often denoted by $||\mathbf{a}||$ or $|\mathbf{a}|$.

When calculating the distance between two points in the plane, it is frequently convenient to use the fact that

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}), \quad \text{for } \mathbf{a}, \mathbf{b} \in \mathbb{R}^2.$$

Exercise 3.1

Prove that, for each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$,

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}).$$

These relationships have a geometric interpretation. When the line segment between $\mathbf{a}$ and $\mathbf{b}$ is translated so that one of its ends is at the origin, the distance between its endpoints is unchanged, as Figure 3.4 illustrates.

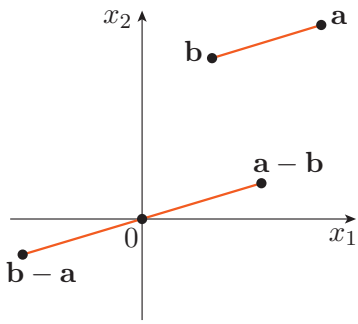


Figure 3.4 Translating a line segment leaves the distance between its endpoints unchanged

Convergence of sequences in the plane

Now that we know how to measure distances in the plane, we can define convergence of planar sequences.

Recall the following result from Definition 1.17.

For a real sequence (a_n), $a_n \to l$ as $n \to \infty$ if, and only if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n - l| < \varepsilon \quad \text{whenever } n > N.$$

Thus if we replace $|a_n - l|$ with $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})$ for $\mathbf{a}_n$ and $\boldsymbol{\ell}$ points in $\mathbb{R}^2$, then we have a possible definition for convergence of sequences in the plane: we could say that a sequence $(\mathbf{a}_n)$ in $\mathbb{R}^2$ converges to a point $\boldsymbol{\ell} \in \mathbb{R}^2$ if for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) < \varepsilon$ whenever $n > N$. But now consider the sequence (b_n) given by $b_n = d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})$ for each $n \in \mathbb{N}$. This is a sequence of (non-negative) real numbers that converges to zero as $n \to \infty$ – that is, it is a null sequence – and so this is an alternative way of saying the same thing.

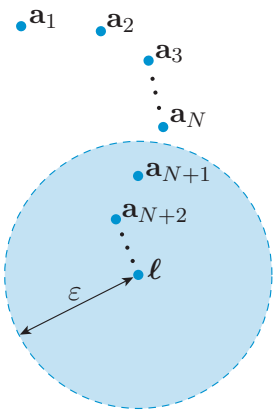


Figure 3.5 Convergence of $(\mathbf{a}_n)$

See Exercise 3.1.

Definition 3.4 *Convergence in* $\mathbb{R}^2$

A sequence $(\mathbf{a}_n)$ in $\mathbb{R}^2$ **converges** to $\boldsymbol{\ell} \in \mathbb{R}^2$ if $\left(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})\right)$ is a (real) null sequence.

We say that $\boldsymbol{\ell}$ is the **limit** of the sequence $(\mathbf{a}_n)$ and write $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$, $\lim_{n\to\infty} \mathbf{a}_n = \boldsymbol{\ell}$ or simply $\mathbf{a}_n \to \boldsymbol{\ell}$.

A sequence in the plane that does not converge to any point in $\mathbb{R}^2$ is said to be **divergent**.

Figure 3.5 illustrates a sequence $(\mathbf{a}_n)$ converging to $\boldsymbol{\ell}$, with $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) < \varepsilon$ for all $n > N$.

Remarks

1. Since $d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) = d^{(2)}(\mathbf{a}_n - \boldsymbol{\ell}, \mathbf{0})$, the statement '$(\mathbf{a}_n)$ converges to $\boldsymbol{\ell}$' is equivalent to '$(\mathbf{a}_n - \boldsymbol{\ell})$ is a sequence that converges to $\mathbf{0}$'.
2. By use of the combination rules for null sequences (in $\mathbb{R}$), we see that

   $$d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}) \to 0 \text{ if, and only if, } d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})^2 \to 0$$

   and so to show that $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$, it is enough to show that

   $$(a_{1,n} - l_1)^2 + (a_{2,n} - l_2)^2 \to 0 \text{ as } n \to \infty.$$

Worked Exercise 3.5

Let $(\mathbf{a}_n)$ be the sequence in the plane defined by $\mathbf{a}_n = \left(\frac{1}{n^2}, 5 + \frac{1}{n}\right)$. Show that $(\mathbf{a}_n)$ converges to $(0, 5)$.

Solution

It is enough to show that $(d^{(2)}(\mathbf{a}_n, (0,5))^2)$ is a (real) null sequence.

For $n \in \mathbb{N}$, we have

$$\begin{aligned} d^{(2)}\left(\left(\tfrac{1}{n^2}, 5 + \tfrac{1}{n}\right), (0,5)\right)^2 &= \left(\tfrac{1}{n^2}\right)^2 + \left(\tfrac{1}{n}\right)^2 \\ &= \tfrac{1}{n^4} + \tfrac{1}{n^2}, \end{aligned}$$

a sum of basic null sequences. We conclude (from the combination rules for null sequences) that $(d^{(2)}(\mathbf{a}_n, (0,5))^2)$ is a (real) null sequence.

Hence $\mathbf{a}_n \to (0, 5)$ as $n \to \infty$.

Worked Exercise 3.6

Let $(\mathbf{b}_n)$ be the sequence in the plane defined by $\mathbf{b}_n = (\frac{1}{n}, n)$. Show that $(\mathbf{b}_n)$ is a divergent sequence.

Solution

Let $\boldsymbol{\ell} = (l_1, l_2)$ be a possible point in the plane that $(\boldsymbol{\ell}_n)$ converges to and notice that

$$d^{(2)}(\mathbf{b}_n, \boldsymbol{\ell})^2 = (\tfrac{1}{n} - \ell_1)^2 + (n - \ell_2)^2 \geq (n - \ell_2)^2.$$

But if $N = \text{int}(|\ell_2|) + 1$, then for $n > N$, $(n - \ell_2)^2 > 1$ and so $d^{(2)}(\mathbf{b}_n, \boldsymbol{\ell})^2$ is not a null sequence.

Hence $(\mathbf{b}_n)$ is a divergent sequence.

Exercise 3.2

Let $(\mathbf{b}_n)$ be the sequence defined by $\mathbf{b}_n = \left(\frac{3}{n} + 1, 4 - \frac{2}{n^3}\right)$. Prove that $(\mathbf{b}_n)$ converges, and find its limit.

In the exercises above, you may notice a connection between the convergence of a sequence of points in the plane and the convergence of the individual coordinates of the terms in the sequence. For example, in Worked Exercise 3.5 the sequence given by $\mathbf{a}_n = \left(\frac{1}{n^2}, 5 + \frac{1}{n}\right)$ converges to $(0, 5)$, while the coordinate sequences $\left(\frac{1}{n^2}\right)$ and $\left(5 + \frac{1}{n}\right)$ converge to 0 and 5 respectively. The next theorem shows that this holds true in general.

Theorem 3.7

Let $(\mathbf{a}_n) = ((a_{1,n}, a_{2,n}))$ be a sequence in $\mathbb{R}^2$ and let $\boldsymbol{\ell} = (\ell_1, \ell_2) \in \mathbb{R}^2$. Then

$\mathbf{a}_n \to \boldsymbol{\ell}$ if, and only if, both $a_{1,n} \to \ell_1$ and $a_{2,n} \to \ell_2$ as $n \to \infty$.

Proof Let us first suppose that $(\mathbf{a}_n)$ converges to $\boldsymbol{\ell}$ and prove that $a_{1,n} \to \ell_1$ and $a_{2,n} \to \ell_2$. This means that we need to show that the real sequences $(a_{1,n} - \ell_1)_{n\in\mathbb{N}}$ and $(a_{2,n} - \ell_2)_{n\in\mathbb{N}}$ are null. We estimate

$$0 \leq |a_{1,n} - \ell_1| \leq \sqrt{(a_{1,n} - \ell_1)^2 + (a_{2,n} - \ell_2)^2} = d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}).$$

We are assuming that $(\mathbf{a}_n)$ converges to $\boldsymbol{\ell}$, and so the real sequence $(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell}))_{n\in\mathbb{N}}$ must be null. Thus it follows from the Squeeze Rule that $(a_{1,n} - \ell_1)_{n\in\mathbb{N}}$ is null.

Hence $a_{1,n} \to \ell_1$. The proof that $a_{2,n} \to \ell_2$ as $n \to \infty$ is similar.

To complete the proof, we now suppose that $a_{1,n} \to \ell_1$ and $a_{2,n} \to \ell_2$, and show that this implies $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$.

So, given that $(a_{1,n} - \ell_1)$ and $(a_{2,n} - \ell_2)$ are null sequences, we use the Product and Sum Rules to deduce that $\left((a_{1,n} - \ell_1)^2 + (a_{2,n} - \ell_2)^2\right)$ is a null sequence of non-negative real numbers. Then the Power Rule for null sequences (see Theorem 1.14) gives that the sequence

$$(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})) = \left(\sqrt{(a_{1,n} - \ell_1)^2 + (a_{2,n} - \ell_2)^2}\right)$$

is null. We conclude that $\mathbf{a}_n$ converges to $\boldsymbol{\ell}$ as $n \to \infty$, as required. ■

Remarks

1. Thus a sequence $(\mathbf{a}_n)$ in the plane is divergent if, and only if, at least one of the component sequences $(a_{1,n})$ and $(a_{2,n})$ is divergent.
2. This result enables us to prove combination rules for sequences in the plane based upon those for real sequences.

Exercise 3.3

(a) Let $(\mathbf{a}_n)$ be the sequence defined by $\mathbf{a}_n = \left(\frac{n^2}{3^n}, \frac{3n! + 2^n}{n^{50} + n!}\right)$. Show that $(\mathbf{a}_n)$ converges, and find its limit.

(b) Suppose that $(\mathbf{a}_n)$ and $(\mathbf{b}_n)$ are sequences in the plane that converge to $\boldsymbol{\ell}$ and $\mathbf{m}$ respectively, as $n \to \infty$. Show that the sequence $(\mathbf{c}_n)$ given by $\mathbf{c}_n = \mathbf{a}_n + \mathbf{b}_n$ converges to $\boldsymbol{\ell} + \mathbf{m}$ as $n \to \infty$.

We could, in a similar way, prove other combination rules for sequences in the plane: we avoid the temptation to do so.

3.3 Continuity on the plane

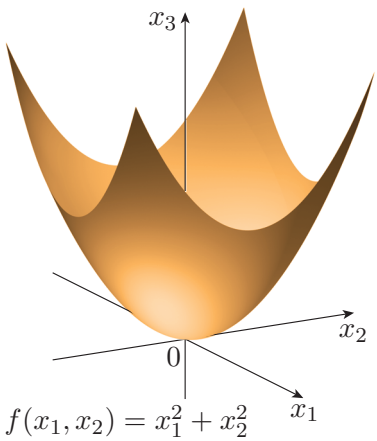


Figure 3.6 Graph of $f(x_1, x_2) = x_1^2 + x_2^2$

In the final part of this section (and chapter), we pull together what we have learned to define continuity for functions from the plane to the real line.

Consider the graph of the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = x_1^2 + x_2^2,$$

shown in Figure 3.6. There are no obvious jumps or gaps, and no regions where the function appears to behave oddly.

It certainly seems reasonable to expect that for any reasonable definition of continuity, this function should be continuous on $\mathbb{R}^2$. However, the definition of continuity given in Section 2 applies only to functions from $\mathbb{R}$ to $\mathbb{R}$. To prove continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$, such as this one, we need a new definition, which should be related in a natural way to the one for functions from $\mathbb{R}$ to $\mathbb{R}$. To show how this can be done, we first give a reminder of the definition of continuity from $\mathbb{R}$ to $\mathbb{R}$.

Continuity from $\mathbb{R}$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}$. A function $f\colon A \to \mathbb{R}$ is **continuous** at $a \in A$ if:

for each sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$,
the sequence $(f(x_n))$ converges to $f(a)$.

We need to change this definition to refer to the fact that the domain of f is a subset of $\mathbb{R}^2$. The changes are fairly straightforward: A becomes a subset of $\mathbb{R}^2$; the point a in $\mathbb{R}$ becomes a point $\mathbf{a}$ in $\mathbb{R}^2$; and in place of real sequences (x_n) we consider sequences $(\mathbf{x}_n)$ in the plane. These changes lead to the following definition.

Definition 3.8 *Continuity at a point in the plane*

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function.

Then f is **continuous** at $\mathbf{a} \in A$ if:

whenever $(\mathbf{x}_n)$ is a sequence in A for which $\mathbf{x}_n \to \mathbf{a}$ as $n \to \infty$,
then the (real) sequence $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$.

If f does not satisfy this condition at $\mathbf{a} \in A$ – that is, there is a sequence $(\mathbf{x}_n)$ in A for which $\mathbf{x}_n \to \mathbf{a}$ as $n \to \infty$ but $f(\mathbf{x}_n)$ does not converge to $f(\mathbf{a})$ – then we say that f is **discontinuous** at $\mathbf{a}$.

To show that f is continuous at $\mathbf{a}$, we must show that $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$ for *every possible choice* of a sequence $(\mathbf{x}_n)$ in A that converges to $\mathbf{a}$.

We extend the definition of continuity at a point to define continuity on a subset of the plane.

Definition 3.9 *Continuity on a subset of the plane*

Let $S \subseteq A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Our first examples of continuous functions from the plane to the real line are very simple but will prove useful later.

Definition 3.10 *Projection functions in the plane*

The two functions $p_1\colon \mathbb{R}^2 \to \mathbb{R}$ and $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$p_1(x_1, x_2) = x_1, \quad p_2(x_1, x_2) = x_2,$$

are known as **projection functions**.

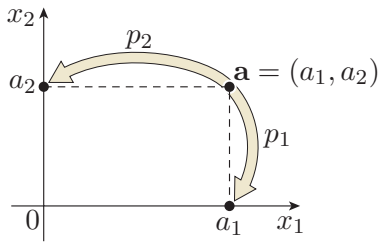


Figure 3.7 The projection functions

These functions *project* each point $\mathbf{a} = (a_1, a_2)$ in $\mathbb{R}^2$ onto the coordinate axes, as Figure 3.7 illustrates.

Worked Exercise 3.11

Prove that the projection functions p_1 and p_2 are continuous on $\mathbb{R}^2$.

Solution

We give the proof for p_1. The proof for p_2 is similar.

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$ be a general point in the plane, and let $(\mathbf{x}_n) = ((x_{1,n}, x_{2,n}))$ be a sequence in $\mathbb{R}^2$ that converges to $\mathbf{a}$. We need to show that $(p_1(\mathbf{x}_n))$ converges to $p_1(\mathbf{a})$.

We have that $p_1(\mathbf{x}_n) = x_{1,n}$ and $p_1(\mathbf{a}) = a_1$.

Since we are assuming that $\mathbf{x}_n \to \mathbf{a}$, we deduce from Theorem 3.7 that $x_{1,n} \to a_1$.

Hence the sequence $(p_1(\mathbf{x}_n))_{n\in\mathbb{N}} = (x_{1,n})_{n\in\mathbb{N}}$ converges to $p_1(\mathbf{a}) = a_1$, and therefore p_1 is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point, p_1 is continuous on $\mathbb{R}^2$.

To show that a function from $\mathbb{R}^2$ to $\mathbb{R}$ is discontinuous, we need to find *just one sequence* $\mathbf{x}_n \to \mathbf{a}$ such that $(f(\mathbf{x}_n))$ does not converge to $f(\mathbf{a})$.

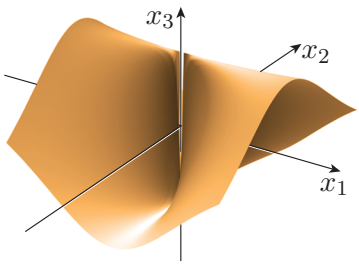


Figure 3.8 Graph of function for Worked Exercise 3.12

Worked Exercise 3.12

Show that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \begin{cases} \dfrac{x_1^2 - x_2^2}{x_1^2 + x_2^2}, & \text{if } \mathbf{x} \neq \mathbf{0}, \\ 0, & \text{if } \mathbf{x} = \mathbf{0}, \end{cases}$$

is discontinuous at $\mathbf{0}$. (The graph of the function is shown in Figure 3.8.)

Solution

Look along the positive x_1-axis, where $x_2 = 0$. There, $f(\mathbf{x}) = f(x_1, 0) = 1$, except at the origin $\mathbf{0}$, where $f(\mathbf{0}) = 0$.

In order to show discontinuity at $\mathbf{0}$, we need only to find a single sequence $\mathbf{x}_n \to \mathbf{0}$ such that $f(\mathbf{x}_n) \not\to f(\mathbf{0}) = 0$ as $n \to \infty$.

Using the behaviour of f along the x_1-axis as our guide, we choose $\mathbf{x}_n = (\frac{1}{n}, 0)$ for $n \in \mathbb{N}$. Then

$$d^{(2)}(\mathbf{x}_n, \mathbf{0}) = \tfrac{1}{n} \to 0 \text{ as } n \to \infty,$$

and so $\mathbf{x}_n \to \mathbf{0}$.

However,

$$f(\mathbf{x}_n) = f(\tfrac{1}{n}, 0) = 1, \quad \text{for each } n \in \mathbb{N},$$

and so $(f(\mathbf{x}_n))$ is just the constant sequence (1), which does not converge to $f(\mathbf{0}) = 0$.

Hence f is discontinuous at $\mathbf{0}$.

Any sequence given by $\mathbf{x}_n = (x_{1,n}, 0)$ with $x_{1,n} \neq 0$ and $x_{1,n} \to 0$ would do.

Along the line $x_1 = x_2$, $f(\mathbf{x})$ is always 0. Although this alone might suggest that f is continuous at $\mathbf{0}$, this is not the case. To be continuous at $\mathbf{0}$, it is necessary that $f(\mathbf{x})$ must approach the value at $f(\mathbf{0})$ *however* $\mathbf{x}$ *approaches* $\mathbf{0}$.

Exercise 3.4

Show that the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(x_1, x_2) = \begin{cases} 0, & \text{if } x_2 \geq x_1, \\ 1, & \text{if } x_2 < x_1, \end{cases}$$

is discontinuous at $(2, 2)$.

The next worked exercise involves a more complicated function and the solution requires some extra effort.

Worked Exercise 3.13

Prove that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x_1, x_2) = x_1^2 + x_2^2$$

is continuous on $\mathbb{R}^2$. (The graph of the function is shown in Figure 3.9.)

Solution

Let $\mathbf{a} \in \mathbb{R}^2$ and let $(\mathbf{x}_n) = ((x_{1,n}, x_{2,n}))$ be a sequence that converges to $\mathbf{a}$.

We prove that $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$ by showing that the sequence of real numbers $(f(\mathbf{x}_n) - f(\mathbf{a}))_{n\in\mathbb{N}}$ is null.

First, we seek an upper bound for $|f(\mathbf{x}_n) - f(\mathbf{a})|$ in terms of $d^{(2)}(\mathbf{x}_n, \mathbf{a})$:

$$\begin{aligned} |f(\mathbf{x}_n) - f(\mathbf{a})| &= |(x_{1,n}^2 + x_{2,n}^2) - (a_1^2 + a_2^2)| \\ &= |(x_{1,n}^2 - a_1^2) + (x_{2,n}^2 - a_2^2)| \\ &\leq |x_{1,n}^2 - a_1^2| + |x_{2,n}^2 - a_2^2| \text{ (by the Triangle Inequality)} \\ &= |x_{1,n} - a_1|\,|x_{1,n} + a_1| + |x_{2,n} - a_2|\,|x_{2,n} + a_2|. \quad (1) \end{aligned}$$

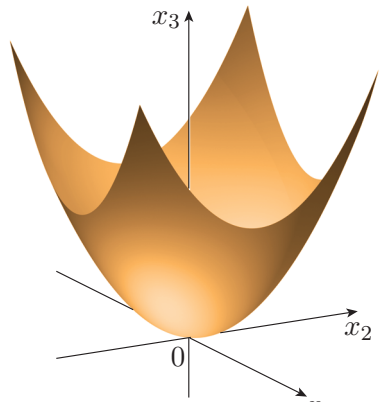


Figure 3.9 Graph of function for Worked Exercise 3.13

But

$$|x_{1,n} - a_1| = \sqrt{(x_{1,n} - a_1)^2}$$
$$\leq \sqrt{(x_{1,n} - a_1)^2 + (x_{2,n} - a_2)^2} = d^{(2)}(\mathbf{x}_n, \mathbf{a}), \quad (2)$$

and, similarly,

$$|x_{2,n} - a_2| \leq d^{(2)}(\mathbf{x}_n, \mathbf{a}). \quad (3)$$

Hence, from equations (1), (2) and (3),

$$|f(\mathbf{x}_n) - f(\mathbf{a})| \leq (|x_{1,n} + a_1| + |x_{2,n} + a_2|)d^{(2)}(\mathbf{x}_n, \mathbf{a}). \quad (4)$$

It still remains to find upper bounds for $|x_{1,n} + a_1|$ and $|x_{2,n} + a_2|$. To do this for $|x_{1,n} + a_1|$, we rewrite it as $|(x_{1,n} - a_1) + 2a_1|$ and use the Triangle Inequality and equation (2) to obtain

$$|x_{1,n} + a_1| = |(x_{1,n} - a_1) + 2a_1| \leq |x_{1,n} - a_1| + 2|a_1|$$
$$\leq d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1|. \quad (5)$$

Similarly, for $|x_{2,n} + a_2|$,

$$|x_{2,n} + a_2| \leq d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_2|. \quad (6)$$

Hence, from equations (4), (5) and (6),

$$0 \leq |f(\mathbf{x}_n) - f(\mathbf{a})| \leq \left(2d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1| + 2|a_2|\right) d^{(2)}(\mathbf{x}_n, \mathbf{a}). \quad (7)$$

Since $\mathbf{x}_n \to \mathbf{a}$, the sequence $\left(d^{(2)}(\mathbf{x}_n, \mathbf{a})\right)_{n\in\mathbb{N}}$ is null and so the sequence $\left(2d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1| + 2|a_2|\right)$ converges to $2|a_1| + 2|a_2|$ by the Sum Rule. The Product Rule now gives that

$$\left(2d^{(2)}(\mathbf{x}_n, \mathbf{a}) + 2|a_1| + 2|a_2|\right) d^{(2)}(\mathbf{x}_n, \mathbf{a}) \to (2|a_1| + 2|a_2|) \times 0 = 0.$$

That is, the right-hand side of equation (7) is a null sequence. By applying the Squeeze Rule to equation (7), we deduce that $(f(\mathbf{x}_n) - f(\mathbf{a})) \to 0$ as $n \to \infty$.

Therefore $f(\mathbf{x}_n) \to f(\mathbf{a})$, and so f is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point, f is continuous on $\mathbb{R}^2$.

As Worked Exercise 3.13 illustrates, it would be tedious if we had to work directly from the definition of continuity whenever we had to prove that a function from $\mathbb{R}^2$ to $\mathbb{R}$ is continuous. Fortunately, as with functions from $\mathbb{R} \to \mathbb{R}$, there are combination rules that allow us to combine continuous functions to form new continuous functions. We state them without proof.

Combination rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ be functions that are continuous at $\mathbf{a} \in A$.

Then the following functions are continuous at $\mathbf{a}$:

Sum Rule	$f + g$, defined by $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$
Multiple Rule	λf where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$
Product Rule	fg, defined by $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$
Quotient Rule	f/g, defined by $(f/g)(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$, provided $g(\mathbf{a}) \neq 0$.

These combination rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ follow from the combination rules for convergent real sequences.

Worked Exercise 3.14

Let $\mathbf{u} = (u_1, u_2)$ be a fixed point in $\mathbb{R}^2$. Use the combination rules to prove that the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$g(\mathbf{x}) = u_1x_1 + u_2x_2$$

is continuous on $\mathbb{R}^2$.

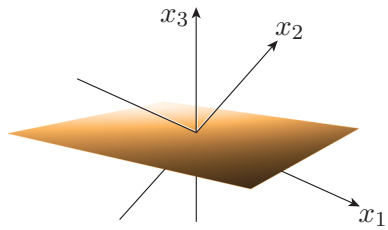
Note that the graph of g is a plane through the origin, as Figure 3.10 illustrates for $\mathbf{u} = (2, -3)$. It is intuitively clear that g is continuous everywhere, but this still needs proof.

Solution

The key to proving that this function is continuous is to recognise that g can be written in terms of the projection functions p_1 and p_2. In fact,

$$g(\mathbf{x}) = u_1p_1(\mathbf{x}) + u_2p_2(\mathbf{x}).$$

Since both p_1 and p_2 are continuous on $\mathbb{R}^2$ and since $\mathbf{u} = (u_1, u_2)$ is fixed, the Multiple Rule guarantees that u_1p_1 and u_2p_2 are both continuous functions on $\mathbb{R}^2$. We can now apply the Sum Rule to u_1p_1 and u_2p_2 to deduce that $g = u_1p_1 + u_2p_2$ is continuous on $\mathbb{R}^2$.



$g(x_1, x_2) = 2x_1 - 3x_2$

Figure 3.10 Graph of $g(\mathbf{x})$ for $\mathbf{u} = (2, -3)$

The result of Worked Exercise 3.14 applies for any $\mathbf{u}$; so it tells us, for example, that the function $f(\mathbf{x}) = 2x_1 - 3x_2$, which is $g(\mathbf{x})$ with $\mathbf{u} = (2, -3)$, is continuous on $\mathbb{R}^2$.

A similar argument works for other functions: for example, an alternative way to verify the continuity of the function f in Worked Exercise 3.13 is to notice that if $f(x_1, x_2) = x_1^2 + x_2^2$, then $f(\mathbf{x}) = (p_1(\mathbf{x}))^2 + (p_2(\mathbf{x}))^2$ and so, since (by Worked Exercise 3.11) p_1 and p_2 are continuous on $\mathbb{R}^2$, we can use the Product Rule (applied to $p_1 \times p_1$ and $p_2 \times p_2$) followed by the Sum Rule to deduce that f is continuous on $\mathbb{R}^2$.

There is also a Restriction Rule for functions from $\mathbb{R}^2$ to $\mathbb{R}$, which allows us to restrict the domain of a continuous function to a smaller set and still have a continuous function.

See Definition 2.7 for the definition of the restricted function.

Proposition 3.15 *Restriction Rule*

Let $A \subseteq \mathbb{R}^2$, let $f\colon A \to \mathbb{R}$ and let $B \subseteq A$. If f is continuous at $\mathbf{b} \in B$, then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous at $\mathbf{b}$.

In particular, if f is continuous on A, then $f|_B$ is continuous on B.

Another useful rule is the one that allows us to compose continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ with continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Again we state it without proof.

Proposition 3.16 *Composition Rule*

Let $A \subseteq \mathbb{R}^2$ and suppose that $f\colon A \to \mathbb{R}$ is continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}$ and suppose $g\colon B \to \mathbb{R}$ is continuous on B. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

Notice that we are *not* composing two functions from $\mathbb{R}^2$ to $\mathbb{R}$. Since the codomain of f is $\mathbb{R}$, and not $\mathbb{R}^2$, the domain of g must be a subset of $\mathbb{R}$. Thus g maps from some subset of $\mathbb{R}$ to $\mathbb{R}$.

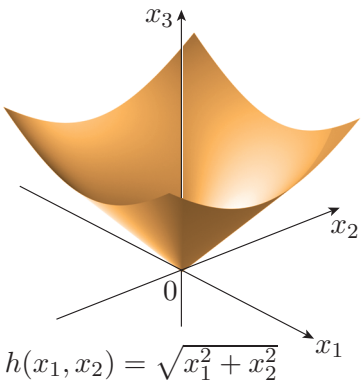


Figure 3.11 Graph of $h(\mathbf{x})$ for Worked Exercise 3.17

Worked Exercise 3.17

Prove that the function $h\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$h(\mathbf{x}) = \sqrt{x_1^2 + x_2^2}$$

is continuous on $\mathbb{R}^2$. (The graph of the function is shown in Figure 3.11.)

Solution

We see that h consists of composing the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by $f(x_1, x_2) = x_1^2 + x_2^2$ with the square-root function. We have already shown that f is continuous on $\mathbb{R}^2$ (Worked Exercise 3.13) and we know that the square-root function is a basic continuous function from $[0, \infty)$ to $\mathbb{R}$. Furthermore, since $f(x_1, x_2) = x_1^2 + x_2^2 \geq 0$ for all real numbers, the image set of f is included in the domain $[0, \infty)$ of the square-root function. Hence the conditions of the Composition Rule are satisfied, so we conclude that h is continuous on $\mathbb{R}^2$.

Exercise 3.5

Use the projection functions and the rules for combining continuous functions to prove that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \sin(x_1 + x_2)$$

is continuous on $\mathbb{R}^2$.

Exercise 3.6

Use the projection functions and the rules for combining continuous functions to prove that the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\mathbf{x}) = e^{-x_1 x_2}$$

is continuous on $\mathbb{R}^2$.

3.4 Properties of distance

Because we wish to generalise the idea of distance to higher dimensions, and to other contexts, we finish this chapter by highlighting certain properties of the usual Euclidean distance.

So far, we have met two ways of defining distance: the distance between points a and b on the real line given by $|b - a|$, and the distance between points $\mathbf{a}$ and $\mathbf{b}$ in the plane given by

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Another definition of distance with which you may be familiar is the great circle distance between points on the surface of a sphere. You will meet other kinds of distance in Chapter 14.

Both of these notions of distance can be thought of as functions of two variables. On the real line, we associate with each pair of real numbers a and b a single non-negative real number $|b - a|$. Similarly, on the plane, we associate with each pair of points (a_1, a_2) and (b_1, b_2) the single non-negative real number

$$\sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Both functions are referred to as *Euclidean distance*, on $\mathbb{R}$ and on $\mathbb{R}^2$ respectively.

Each of these distance functions has three significant properties, the first two of which may seem almost trivial. We draw attention to these three properties because they are the key properties that we will use to define the general notion of a distance function. We concentrate on describing these properties for the Euclidean distance function $d^{(2)}$ on $\mathbb{R}^2$.

The first property is that the distance between two *distinct* points is always positive, and the distance between two points that coincide is 0. This property, which we label (M1), is stated formally as follows.

M stands for *metric* – of which you will see more in the next chapter.

(M1) For each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, we have $d^{(2)}(\mathbf{a}, \mathbf{b}) \geq 0$, with $d^{(2)}(\mathbf{a}, \mathbf{b}) = 0$ if, and only if, $\mathbf{a} = \mathbf{b}$.

The second property is that it does not matter in which direction you measure the distance between two points: the distance from $\mathbf{a}$ to $\mathbf{b}$ is the same as the distance from $\mathbf{b}$ to $\mathbf{a}$. We say that $d^{(2)}$ is *symmetric*.

(M2) For each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, we have $d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{b}, \mathbf{a})$.

Worked Exercise 3.18

Prove that the distance property (M2) holds for $d^{(2)}$.

Solution

This follows from the fact that, for any $p, q \in \mathbb{R}$, $(p-q)^2 = (q-p)^2$.

Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$. Then

$$\begin{aligned} d^{(2)}(\mathbf{a}, \mathbf{b}) &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \\ &= \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2} = d^{(2)}(\mathbf{b}, \mathbf{a}). \end{aligned}$$

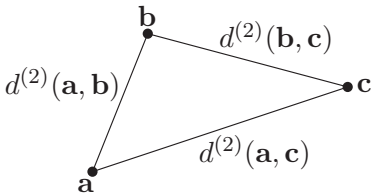


Figure 3.12 The Triangle Inequality

The third important property is the **Triangle Inequality**. This inequality expresses the fact that in a triangle, the length of any one side cannot exceed the sum of the lengths of the other two sides. It can be expressed symbolically as follows, and is illustrated in Figure 3.12.

(M3) For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$, we have $d^{(2)}(\mathbf{a}, \mathbf{c}) \leq d^{(2)}(\mathbf{a}, \mathbf{b}) + d^{(2)}(\mathbf{b}, \mathbf{c})$.

Exercise 3.7

Verify the Triangle Inequality for the points

$$\mathbf{a} = (3, 4), \quad \mathbf{b} = (6, 0), \quad \mathbf{c} = (5, -3).$$

There are several ways to prove the Triangle Inequality for $d^{(2)}$. We will prove a general version for Euclidean distance on $\mathbb{R}^n$ in the next chapter, obtaining the Triangle Inequality for the plane as the special case $n = 2$.

The three properties (M1)–(M3) also apply to the Euclidean distance function $|b - a|$ on $\mathbb{R}$, for which they were demonstrated in Subsection 1.1. They can be summarised as follows.

(M1) For each $a, b \in \mathbb{R}$, $|b - a| \geq 0$, with $|b - a| = 0$ if, and only if, $b = a$.

(M2) For each $a, b \in \mathbb{R}$, $|b - a| = |a - b|$.

(M3) For each $a, b, c \in \mathbb{R}$, $|c - a| \leq |b - a| + |c - b|$.

The properties (M1)–(M3) hold not only for the Euclidean distance functions defined on $\mathbb{R}$ and $\mathbb{R}^2$, but also for Euclidean distance in higher dimensions. In fact, these three properties are the key properties that we expect any reasonable notion of 'distance' to possess. In Chapter 14, we will specify that *any* set endowed with a function satisfying properties (M1)–(M3) is a set for which distance is defined. We will then be in a position to explore related properties such as convergence of sequences and continuity in a wider setting.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of this chapter.

After working through this chapter you should be able to:

(a) State the definition of a function.

(b) Use the modulus function in inequalities and obtain the solution sets of inequalities involving the modulus function.

(c) Use the Triangle Inequality and its variants in $\mathbb{R}$.

(d) Explain the meanings of *sequence*, *null sequence*, *convergent sequence*, *divergent sequence* and *subsequence*.

(e) Test a given real sequence for convergence or divergence.

(f) Use the (sequential) definition of continuity to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous.

(g) Use the rules for continuous functions to determine whether a given function from $\mathbb{R}$ to $\mathbb{R}$ is continuous.

(h) State the Intermediate Value Theorem and use it to show that certain equations have roots.

(i) State the Boundedness and Extreme Value Theorems.

(j) Calculate the distance between any two points in the plane $\mathbb{R}^2$.

(k) Test a given sequence in $\mathbb{R}^2$ for convergence or divergence.

(l) Explain the definition of continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$.

(m) Determine whether a given function from $\mathbb{R}^2$ to $\mathbb{R}$ is continuous.

(n) Understand the Euclidean distance function on $\mathbb{R}^2$, and appreciate its properties.

Solutions to exercises

Solution to Exercise 1.1

(a) This is a function from $\mathbb{Z}$ to $\mathbb{N}$, since the square of an integer is a non-negative integer and so adding 1 to this number always gives a natural number.

(b) This is not a function from $\mathbb{Z}$ to $\mathbb{N}$, since it associates with a given integer a *set* of natural numbers, and this is not an element of the codomain.

(c) This is not a function from $\mathbb{Z}$ to $\mathbb{N}$, since the integer 0 has no well-defined image. In particular, each non-zero integer n is a factor of 0 and so the set of factors of 0 is $\mathbb{Z} - \{0\}$, which is an infinite set.

Solution to Exercise 1.2

Only the rule given in Exercise 1.1(a) defines a function, $f\colon \mathbb{Z} \to \mathbb{N}$ given by $f(m) = m^2 + 1$. For this function,

$$f(\mathbb{N}) = \{f(n) : n \in \mathbb{N}\} = \{n^2 + 1 : n \in \mathbb{N}\} = \{2, 5, 10, 17, 26, \ldots\}$$

and

$$f^{-1}(\{5\}) = \{x \in \mathbb{Z} : f(x) \in \{5\}\} = \{x \in \mathbb{Z} : x^2 + 1 \in \{5\}\} = \{-2, 2\}.$$

Solution to Exercise 1.3

The answers to parts (a) and (b) can be read off immediately from Proposition 1.2.

(a) The solution set is $(-3, 3)$.

(b) The solution set is $\emptyset$.

(c) $|3x| < 3$ if, and only if, $|3|\,|x| < 3$, which is true if, and only if, $3|x| < 3$, which holds if, and only if, $|x| < 1$. Thus $\{x \in \mathbb{R} : |3x| < 3\} = \{x \in \mathbb{R} : |x| < 1\}$, so the solution set is the interval $(-1, 1)$.

Solution to Exercise 1.4

On noting that $|a| = ||a||$, we can read directly from Worked Exercise 1.3 that

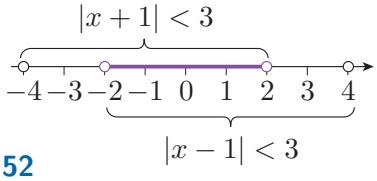
$$|a| = ||a|| \leq |b| \text{ if, and only if, } -|b| \leq |a| \leq |b|,$$

as required.

Solution to Exercise 1.5

(a) By the first result in Proposition 1.5, with $b = 1$ and $c = 3$, the solution set is the open interval $(1 - 3, 1 + 3) = (-2, 4)$.

(b) From part (a), the solution set of $|x - 1| < 3$ is the open interval $(-2, 4)$. Similarly, the solution set of $|x + 1| < 3$ is the open interval $(-4, 2)$. Since both inequalities must hold, we take the intersection of the intervals $(-2, 4)$ and $(-4, 2)$. This is the open interval $(-2, 2)$, as shown.



Solution to Exercise 1.6

We prove the contrapositive: that is, instead of showing that for non-negative numbers $p^2 \leq q^2 \implies p \leq q$, we prove the equivalent statement that for non-negative numbers p and q, if $q < p$, then $q^2 < p^2$.

Suppose that $0 \leq q < p$. Then, since inequalities are preserved when multiplying by a positive real number, $0 \leq q^2 < pq < p^2$. The result follows.

Solution to Exercise 1.7

(a) We use the Triangle Inequality twice.

Let $d = b + c$. Then

$$|a + b + c| = |a + d| \leq |a| + |d|,$$

by the Triangle Inequality. Also,

$$|d| = |b + c| \leq |b| + |c|.$$

Combining these two inequalities gives

$$|a + b + c| \leq |a| + |b| + |c|.$$

(b) The Triangle Inequality states that, for *all* real numbers x and y,

$$|x + y| \leq |x| + |y|.$$

Let $x = a$ and $y = -b$. Then

$$|a - b| = |a + (-b)| \leq |a| + |-b| = |a| + |b|.$$

Solution to Exercise 1.8

For real numbers a and b, we use the Triangle Inequality to find

$$|b| = |(b - a) + a| \leq |b - a| + |a|$$

and

$$|a| = |(a - b) + b| \leq |a - b| + |b| = |b - a| + |b|.$$

Hence

$$|b| - |a| \leq |b - a|$$

and

$$|a| - |b| \leq |b - a|.$$

In particular, since $\big||b| - |a|\big|$ equals one of $|b| - |a|$ or $|a| - |b|$,

$$\big||b| - |a|\big| \leq |b - a|,$$

as required.

Solution to Exercise 1.9

Let $\varepsilon > 0$ be given. Here $|a_n| = \frac{1}{\sqrt{n}}$ and

$$\frac{1}{\sqrt{n}} < \varepsilon \quad \text{if, and only if,} \quad n > \frac{1}{\varepsilon^2}.$$

If we set $N = \text{int}\left(\frac{1}{\varepsilon^2}\right)$ then

$$|a_n| = \frac{1}{\sqrt{n}} < \varepsilon \quad \text{whenever } n > N.$$

Hence (a_n) is null.

Solution to Exercise 1.10

For each n, $|a_n| = |2| = 2$. We take any $\varepsilon \leq 2$, say $\varepsilon = 1$, and so, for each n, $|a_n| = 2 > 1 = \varepsilon$. We deduce that (a_n) is not a null sequence.

Solution to Exercise 1.11

Since (a_n) is null, the Multiple Rule implies that (λa_n) is null. Now $|b_n| \leq \lambda |a_n| = |\lambda a_n|$ (since $\lambda \geq 0$). It follows from the Squeeze Rule that (b_n) is null.

Solution to Exercise 1.12

A sequence (a_n) is divergent if for each real number l, the sequence $(a_n - l)$ is not a null sequence. That is, for each $l \in \mathbb{R}$, there is at least one $\varepsilon > 0$ for which there is *no* natural number N such that $|a_n - l| < \varepsilon$ whenever $n > N$. Alternatively, whatever the value of l, there exists an $\varepsilon > 0$ such that, no matter what choice of N we make, we can always find a natural number $n > N$ for which $|a_n - l| \geq \varepsilon$.

Solution to Exercise 1.13

Let

$$a_n = \frac{4n^2 + 1000^n}{n! + 7n^3}.$$

Dividing both the numerator and denominator of a_n by the dominant term $n!$ gives

$$a_n = \frac{4(n^2/n!) + 1000^n/n!}{1 + 7(n^3/n!)}.$$

The sequences $\left(\frac{n^2}{n!}\right)$, $\left(\frac{1000^n}{n!}\right)$, $\left(\frac{n^3}{n!}\right)$ are all basic null sequences and any constant sequence converges to its constant value. Hence, by the combination rules for convergent sequences,

$$a_n \to \frac{(4 \times 0) + 0}{1 + (7 \times 0)} = 0 \quad \text{as } n \to \infty.$$

Thus (a_n) converges to 0, i.e. it is null.

Solution to Exercise 1.14

Write $a_n = 2^n$ and let $n_k = k^2$ for each $k \in \mathbb{N}$. Then (n_k) is a strictly increasing sequence of positive integers, and $a_{n_k} = 2^{k^2}$. Thus $(a_{n_k}) = (2^{k^2})$ is a subsequence of $(a_n) = (2^n)$.

Solution to Exercise 1.15

Let

$$a_n = (-1)^n \frac{1-n}{n}.$$

Consider the even and odd subsequences (a_{2k}) and (a_{2k-1}).

For (a_{2k}), we have

$$a_{2k} = (-1)^{2k} \frac{1-2k}{2k} = \frac{1-2k}{2k}.$$

Dividing both the numerator and denominator by the dominant term k gives

$$a_{2k} = \frac{\frac{1}{k} - 2}{2}.$$

Now $(\frac{1}{k})$ is a basic null sequence, so the combination rules for convergent sequences imply that

$$a_{2k} \to \frac{0-2}{2} = -1 \quad \text{as } k \to \infty.$$

For (a_{2k-1}), we have

$$\begin{aligned} a_{2k-1} &= (-1)^{2k-1} \frac{1-(2k-1)}{2k-1} \\ &= (-1)\frac{2-2k}{2k-1} = \frac{2k-2}{2k-1}. \end{aligned}$$

Dividing both the numerator and denominator by the dominant term k gives

$$a_{2k-1} = \frac{2 - \frac{2}{k}}{2 - \frac{1}{k}}.$$

Now $(\frac{1}{k})$ is a basic null sequence and so, by the Multiple Rule for null sequences (with $\lambda = 2$), $(\frac{2}{k})$ is a null sequence. Hence the combination rules for convergent sequences imply that

$$a_{2k-1} \to \frac{2-0}{2-0} = 1 \quad \text{as} \quad k \to \infty.$$

Since the even and odd subsequences have different limits, we deduce from Proposition 1.25 that (a_n) is divergent.

Solution to Exercise 1.16

(a) This is a convergent sequence. To prove this, we use the combination rules for convergent sequences together with our list of basic null sequences.

The sequences $\left(\frac{n^3}{n!}\right)$ and $\left(\frac{1}{n^2}\right)$ are basic null sequences. Hence by the Multiple Rule, so are $\left(\frac{3n^3}{n!}\right)$ and $\left(\frac{7}{n^2}\right)$, and so, by the Sum Rule, so is $\left(\frac{3n^3}{n!}+\frac{7}{n^2}\right)$. Thus, the limit is 0.

(b) This sequence does not converge. We have

$$(n+1)^2 - n^2 = 2n+1.$$

Thus $((n+1)^2 - n^2) = (2n+1)$, which is divergent. To see that $(2n+1)$ is divergent, we show directly that if $a \in \mathbb{R}$, then $2n+1 \not\to a$ as $n \to \infty$.

If $\varepsilon = 1$ and $N \in \mathbb{N}$ is chosen so that $N \geq \frac{|a|}{2}$, then for $n > N$,

$$\begin{aligned}|(2n+1) - a| &\geq 2n + 1 - |a| > 2N + 1 - |a| \\ &\geq 1 = \varepsilon.\end{aligned}$$

(We have used the Reverse Triangle Inequality here.) We conclude that $(2n+1)$ does not converge to a, and so, as a is arbitrary, does not converge to anything.

(c) This is a convergent sequence. We first observe that, for each $n \in \mathbb{N}$,

This is a 'trick' that is worth knowing.

$$\begin{aligned}\sqrt{n+1} - \sqrt{n} &= \frac{(n+1)-n}{\sqrt{n+1}+\sqrt{n}} \\ &= \frac{1}{\sqrt{n+1}+\sqrt{n}}.\end{aligned}$$

Hence

$$|\sqrt{n+1} - \sqrt{n}| = \left|\frac{1}{\sqrt{n+1}+\sqrt{n}}\right| \leq \frac{1}{\sqrt{n}},$$

which is a basic null sequence. We deduce from the Squeeze Rule that $(\sqrt{n+1} - \sqrt{n})$ is null.

Solution to Exercise 2.1

Let a be a general point in $\mathbb{R}$. We show that if (x_n) is a sequence in $\mathbb{R}$ that converges to a, then the sequence $(f(x_n))$ converges to $f(a) = a^3 + 3$.

So let (x_n) be a sequence (in $\mathbb{R}$) that converges to a. Then

$$f(x_n) = x_n^3 + 3 = (x_n \times x_n \times x_n) + 3.$$

Thus, since constant sequences converge to their constant value, we apply the Product Rule for convergent sequences (twice) together with the Sum Rule for convergent sequences and deduce that

$$f(x_n) \to (a \times a \times a) + 3 = a^3 + 3 = f(a) \text{ as } n \to \infty.$$

Hence, $f(x_n) \to f(a)$ as $x_n \to a$, and so f is continuous at a.

Since a is an arbitrary point of $\mathbb{R}$, we conclude that f is continuous on $\mathbb{R}$.

Solution to Exercise 2.2

Let $a \neq 0$. There are two possibilities: either $a > 0$ or $a < 0$.

If $a > 0$, then $f_2(a) = 1$. Suppose that (x_n) is a sequence in $\mathbb{R}$ for which $x_n \to a$ as $n \to \infty$.

Then we can find $N \in \mathbb{N}$ such that $x_n > 0$ for $n > N$ (for example by taking N so that, for $n > N$, $|x_n - a| < |a|$ – ensuring the distance of x_n from a is less than the distance of a from 0).

Thus for $n > N$, $f_2(x_n) = f(a) = 1$, so $f_2(x_n) \to 1 = f_2(a)$ as $n \to \infty$. Hence f_2 is continuous at a.

For $a < 0$, we proceed in a similar manner. Upon noting that $f_2(a) = 0$, we suppose (x_n) is a sequence in $\mathbb{R}$ that converges to a.

Then we can find $N \in \mathbb{N}$ such that $x_n < 0$ for $n > N$, so that $f_2(x_n) = f_2(a) = 0$ for $n > N$. Hence $f_2(x_n) \to 0 = f_2(a)$ as $n \to \infty$, and f_2 is continuous at a.

Since we have covered all possible values of $a \neq 0$, we conclude that f is continuous on $\mathbb{R} - \{0\}$.

Solution to Exercise 2.3

We need to show that for any *null* sequence (x_n), $f(x_n) \to f(0) = 0$.

So suppose (x_n) is null. Then, for each n, *either* $x_n \neq 0$ and $|f(x_n)| = |x_n \cos(\pi/x_n)| = |x_n|\,|\cos(\pi/x_n)|$, *or* $x_n = 0$ and $f(x_n) = 0$.

Now $|\cos a| \leq 1$ for all real numbers a and so, when $x_n \neq 0$,

$$|f(x_n)| = |x_n|\,|\cos(\pi/x_n)| \leq |x_n|.$$

Whereas if $x_n = 0$, then

$$|f(x_n)| = 0 \leq |x_n| = 0.$$

Hence $|f(x_n)| \leq |x_n|$ for each $n \in \mathbb{N}$.

But (x_n) is a null sequence and so, by the Squeeze Rule, $|f(x_n)| \to 0$ as $n \to \infty$. Thus $f(x_n) \to f(0) = 0$ as $x_n \to 0$.

Hence f is continuous at 0.

Solution to Exercise 2.4

(a) For $x \neq 1$,

$$f(x) = \frac{x^2 - 1}{x - 1} = \frac{(x+1)(x-1)}{x-1} = x + 1$$

and so if (x_n) is a sequence in $\mathbb{R} - \{1\}$, then

$$f(x_n) = x_n + 1.$$

So, if $x_n \to 1$ as $n \to \infty$, then the Sum Rule implies that

$$f(x_n) = x_n + 1 \to 1 + 1 = 2 \text{ as } n \to \infty.$$

Hence $(f(x_n))$ is a convergent sequence with limit 2.

(b) To show that g is continuous at $a = 1$, suppose that $x_n \to 1$ as $n \to \infty$. Then for any given n, either $x_n = 1$ and so $g(x_n) = 2$, or $x_n \neq 1$ and $g(x_n) = f(x_n) = x_n + 1$. In either case,

$$|g(x_n) - g(1)| = |g(x_n) - 2| \leq |x_n - 1| \to 0 \quad \text{as } n \to \infty.$$

We deduce that $g(x_n) \to g(1)$, and hence g is continuous at $a = 1$.

In fact, for $x \neq 1$, $f(x) = x + 1$ and so it is no surprise that g is continuous at 1.

Solution to Exercise 2.5

The functions $x \mapsto e^x$, $x \mapsto -4x^3$ and $x \mapsto \sin x$ are basic continuous functions with domain $\mathbb{R}$. The function $x \mapsto 1/x$ is a basic continuous function with domain $\mathbb{R} - \{0\}$. Hence, by the Restriction Rule, $x \mapsto 1/x$ and $x \mapsto -4x^3$ are continuous on $(0, \infty)$. Also, by the Composition Rule, $x \mapsto \sin(1/x)$ and $x \mapsto e^{-4x^3}$ are continuous on $(0, \infty)$. It follows from the Multiple Rule that $x \mapsto 2\sin(1/x)$ and $x \mapsto -3e^{-4x^3}$ are continuous on $(0, \infty)$. Finally, by the Sum Rule, f is continuous on $(0, \infty)$.

Solution to Exercise 2.6

We are given that f is continuous on $\mathbb{R}$. The function $x \mapsto |x|$ is a basic continuous function on $\mathbb{R}$. Hence, by the Composition Rule, the function $|f|$ is continuous on $\mathbb{R}$.

Solution to Exercise 2.7

We first show that $g\colon [0,1] \to \mathbb{R}$ given by $g(x) = f(x) - x$ is continuous. Since $x \mapsto -x$ is a basic continuous function on $\mathbb{R}$, it is continuous on $[0,1]$ (by the Restriction Rule). Since f is continuous on $[0,1]$, so is $g(x)$ (by the Sum Rule).

Now notice that, since $f(0) \geq 0$,

$$g(0) = f(0) - 0 \geq 0$$

and, since $f(1) \leq 1$,

$$g(1) = f(1) - 1 \leq 1 - 1 \leq 0.$$

Hence *either* $g(0) = g(1) = 0$, implying that $f(0) = 0$ and $f(1) = 1$ (and so 0 and 1 are both fixed points of f), *or* $g(0) > g(1)$ and we can use the Intermediate Value Theorem to deduce that there is a number $k \in [0,1]$ for which $g(k) = 0$. But if $g(k) = 0$, then $f(k) - k = 0$, and so $f(k) = k$.

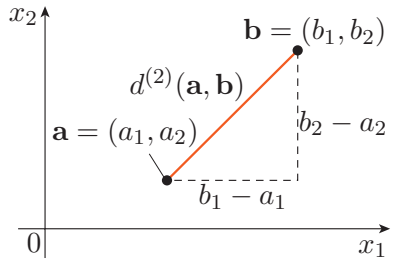
Solution to Exercise 2.8

The function f is the restriction of a basic continuous function, and so is continuous on $[-1,1] - \{0\}$. However, the set $[-1,1] - \{0\}$ is *not* a closed interval, and so the hypotheses of the Boundedness Theorem and the Extreme Value Theorem are not satisfied. So we cannot expect the conclusions of the Boundedness and Extreme Value Theorems to hold – and indeed, they do not: f is bounded neither above nor below, and so certainly cannot attain its bounds.

Solution to Exercise 3.1

Let $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$, as shown. Then

$$\mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2),$$
$$\mathbf{b} - \mathbf{a} = (b_1 - a_1, b_2 - a_2).$$



Thus,

$$\begin{aligned} &d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) \\ &\quad = \sqrt{(0 - (a_1 - b_1))^2 + (0 - (a_2 - b_2))^2} \\ &\quad = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}. \end{aligned}$$

Similarly,

$$\begin{aligned} &d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0}) \\ &\quad = \sqrt{(0 - (b_1 - a_1))^2 + (0 - (b_2 - a_2))^2} \\ &\quad = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}. \end{aligned}$$

Also

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Thus, $d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(2)}(\mathbf{b} - \mathbf{a}, \mathbf{0})$.

Solution to Exercise 3.2

The terms of $(\mathbf{b}_n)$ appear to get near to $(1,4)$ as n gets large, and so we show that $(d^{(2)}(\mathbf{b}_n,(1,4))^2)$ is a real null sequence. For $n \in \mathbb{N}$, we have

$$\begin{aligned} d^{(2)}\left(\left(\tfrac{3}{n}+1, 4-\tfrac{2}{n^3}\right),(1,4)\right)^2 &= \left(\tfrac{3}{n}\right)^2 + \left(-\tfrac{2}{n^3}\right)^2 \\ &= \tfrac{9}{n^2} + \tfrac{4}{n^6}, \end{aligned}$$

a sum of multiples of basic null sequences. Hence $(d^{(2)}(\mathbf{b}_n,(1,4))^2)$ is a null sequence.

We conclude that $(\mathbf{b}_n)$ converges to the limit $(1,4)$.

Solution to Exercise 3.3

(a) By Theorem 3.7, it is enough to show that the coordinate sequences $\left(\dfrac{n^2}{3^n}\right)$ and $\left(\dfrac{3n!+2^n}{n^{50}+n!}\right)$ both converge. We note that $\left(\dfrac{n^2}{3^n}\right)$ is a basic null sequence, while in Worked Exercise 1.20(b) we showed that $\left(\dfrac{3n!+2^n}{n^{50}+n!}\right)$ converges to 3.

Therefore, using Theorem 3.7, we find that $\mathbf{a}_n \to (0,3)$ as $n \to \infty$.

(b) Since $\mathbf{a}_n \to \boldsymbol{\ell}$ and $\mathbf{b}_n \to \mathbf{m}$ as $n \to \infty$, it follows from Theorem 3.7 that

$$a_{1,n} \to l_1,\ a_{2,n} \to l_2,\ b_{1,n} \to m_1 \text{ and } b_{2,n} \to m_2 \text{ as } n \to \infty.$$

Hence, by the Sum Rule for real sequences,

$$c_{1,n} = a_{1,n} + b_{1,n} \to l_1 + m_1$$

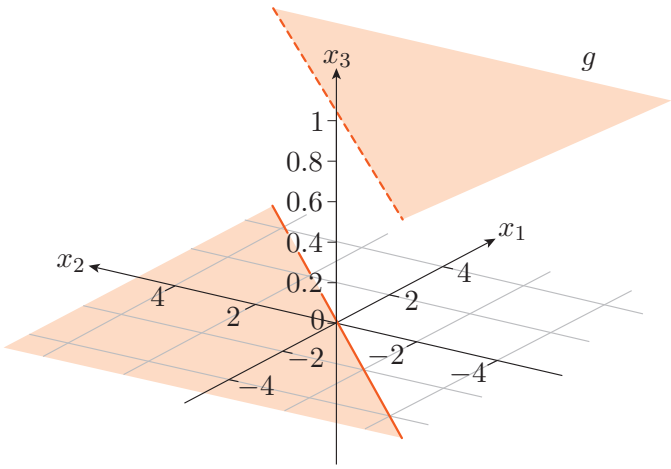
and

$$c_{2,n} = a_{2,n} + b_{2,n} \to l_2 + m_2$$

as $n \to \infty$. We now use Theorem 3.7 again to deduce that $\mathbf{c}_n \to \boldsymbol{\ell} + \mathbf{m}$ as $n \to \infty$.

Solution to Exercise 3.4

The function g (shown in the figure) appears to be discontinuous along the line $x_1 = x_2$: the function values on one side of the line are all equal to 1, and on the other side (and on the line itself) are equal to 0. There are points (x_1, x_2) arbitrarily close to $(2, 2)$ for which $g(x_1, x_2) = 0$, and other points for which $g(x_1, x_2) = 1$.



To prove that g is discontinuous at $(2, 2)$, we need only to find one sequence $\mathbf{x}_n \to (2, 2)$ such that $g(\mathbf{x}_n) \not\to g(2, 2) = 0$.

To do this we choose a sequence that approaches $(2, 2)$ from the $x_1 > x_2$ side, such as $\mathbf{x}_n = (2 + \frac{1}{n}, 2 - \frac{1}{n})$ for $n \in \mathbb{N}$. Then

$$d^{(2)}(\mathbf{x}_n, (2, 2)) = \frac{\sqrt{2}}{n} \to 0 \text{ as } n \to \infty,$$

and so $\mathbf{x}_n \to (2, 2)$ as $n \to \infty$. However,

$$g(\mathbf{x}_n) = g(2 + \tfrac{1}{n}, 2 - \tfrac{1}{n}) = 1, \quad \text{for each } n \in \mathbb{N},$$

and so the sequence $(g(\mathbf{x}_n))$ does not converge to $g(2, 2) = 0$ as $n \to \infty$.

Hence g is discontinuous at $(2, 2)$.

Solution to Exercise 3.5

The function f consists of adding the variables x_1 and x_2 and then taking the sine of the result. The variables x_1 and x_2 are given by the projection functions $p_1(\mathbf{x}) = x_1$ and $p_2(\mathbf{x}) = x_2$, both of which are continuous. Hence, by the Sum Rule, so is $p_1(\mathbf{x}) + p_2(\mathbf{x}) = x_1 + x_2$. This function has image set $\mathbb{R}$, which is the domain of the sine function. The sine function is a basic continuous function. Therefore, by the Composition Rule, we can compose it with $p_1(\mathbf{x}) + p_2(\mathbf{x})$ to give the continuous function $\sin(p_1(\mathbf{x}) + p_2(\mathbf{x})) = \sin(x_1 + x_2) = f(\mathbf{x})$. Thus f is continuous on $\mathbb{R}^2$.

Solution to Exercise 3.6

The variables x_1 and x_2 are given by the projection functions p_1 and p_2, both of which are continuous. Hence so is $-p_1(\mathbf{x})p_2(\mathbf{x}) = -x_1x_2$, by the Multiple and Product Rules. This function has image set $\mathbb{R}$, which is the domain of the function exp. Since exp is a basic continuous function, so is $\exp(-x_1x_2) = e^{-x_1x_2} = f(\mathbf{x})$, by the Composition Rule. Thus f is continuous on $\mathbb{R}^2$.

Solution to Exercise 3.7

$$
\begin{aligned}
&d^{(2)}(\mathbf{a},\mathbf{b}) + d^{(2)}(\mathbf{b},\mathbf{c}) \\
&\quad = \sqrt{(6-3)^2 + (0-4)^2} + \sqrt{(5-6)^2 + (-3-0)^2} \\
&\quad = 5 + \sqrt{10} > 5 + \sqrt{9} = 8;
\end{aligned}
$$

$$
d^{(2)}(\mathbf{a},\mathbf{c}) = \sqrt{(5-3)^2 + (-3-4)^2} = \sqrt{53} < \sqrt{64} = 8.
$$

Hence, $d^{(2)}(\mathbf{a},\mathbf{c}) \leq d^{(2)}(\mathbf{a},\mathbf{b}) + d^{(2)}(\mathbf{b},\mathbf{c})$.

Chapter 14

Metric spaces and continuity 1

Introduction

In Chapter 13, *Distance and continuity*, we saw how the sequential definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ can be used as a basis for a definition of continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$. In this chapter we continue our journey towards understanding continuity in a general setting by finding a definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, which is inspired by our earlier definition of continuity for functions from $\mathbb{R}^2$ to $\mathbb{R}$. The *Euclidean distance* between pairs of points in $\mathbb{R}^n$ (and $\mathbb{R}^m$) plays a key role in this definition. In Chapter 13 we found that the Euclidean distance function for the plane has three properties: non-negativity, symmetry and the Triangle Inequality. In a similar way, we find in this chapter that the Euclidean distance on $\mathbb{R}^n$ satisfies the same three properties. Our creative leap is to use this observation to raise these properties to the status of axioms and to say that any function that satisfies them is a well-defined distance function, known as a *metric*.

In Section 2, we start to explore the implications of this definition and see how the notion of a metric allows our definitions of convergence and continuity in Euclidean spaces to be extended naturally to other spaces. We also see how the concept of a metric leads to different ways of measuring 'distance' on the same underlying set of points.

Chapter guide

Section 1 of this chapter to some extent repeats the work of the previous chapter, but in the context of $\mathbb{R}^n$ rather than the plane. Apart from some extra notational complexity, you will already be familiar with the ideas.

The key idea in the chapter is the decision to use three particular properties of the Euclidean distance as the basis for defining what we mean by a general distance function, a *metric*. This process of using the properties of the (reasonably) familiar as the basis of a definition for a whole class of objects can be challenging when first encountered, and it is likely that you will need to spend a significant proportion of your time studying Section 2.

Sections 3 and 4 are very brief. Section 3 develops the idea of sequences and convergence in metric spaces, before Section 4 extends the definition of continuity from Euclidean spaces to metric spaces.

1 Continuity of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

In this section we continue our process of abstracting the ideas of distance and the continuity of functions. Having dealt with distance and continuity in one and two dimensions, we now look at them in higher dimensions; in particular, we examine distance in $\mathbb{R}^n$ (for general n) and the continuity of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

If we look back at our work in the last chapter, we see that there are several notions that played an important role in reaching a definition of continuity for functions from $\mathbb{R}^2 \to \mathbb{R}$: we needed to be able to measure distance between pairs of points in both $\mathbb{R}$ and $\mathbb{R}^2$, and we also needed a notion of convergence for sequences in both $\mathbb{R}$ and $\mathbb{R}^2$. The methods we adopt for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ are inspired by those we used for functions from $\mathbb{R}^2$ to $\mathbb{R}$ but with some extra notational complication because of the need to track the dimensions of the spaces involved. Since we cannot usually easily visualise distances and functions in such spaces, we rely on algebraic methods to carry through these extensions.

Replacing $\mathbb{R}^n$ by $\mathbb{R}^3$ can sometimes help you to understand general results.

However, before we begin this process, we digress briefly to explain what we mean by A^k when A is a set and $k \in \mathbb{N}$.

Definition 1.1 *Product set*

Let $A_1, A_2, \ldots, A_k$ be sets. The **product set** $A_1 \times A_2 \times \cdots \times A_k$ is defined to be the set consisting of all ordered k-tuples $(a_1, a_2, \ldots, a_k)$ where $a_1 \in A_1$, $a_2 \in A_2$, $\ldots$, $a_k \in A_k$. Thus

$$A_1 \times A_2 \times \cdots \times A_k = \{(a_1, a_2, \ldots, a_k) : a_i \in A_i \text{ for } i = 1, \ldots, k\}.$$

If $A_1 = A_2 = \cdots = A_k = A$, then the product set $A_1 \times A_2 \times \cdots \times A_k$ is written simply as A^k and

$$A^k = \{(a_1, a_2, \ldots, a_k) : a_i \in A \text{ for } i = 1, \ldots, k\}.$$

The product set is sometimes called the *Cartesian product*.

Remarks

1. The term *product* is used for this set because, when $A_1, A_2, \ldots, A_k$ are finite sets with $m_1, m_2, \ldots, m_k$ elements respectively, the set $A_1 \times A_2 \times \cdots \times A_k$ has $m_1 m_2 \cdots m_k$ elements. For example, if $A_1 = \{1, 2\}$ and $A_2 = \{1, 4, 9\}$, then

   $$A_1 \times A_2 = \{(1,1), (1,4), (1,9), (2,1), (2,4), (2,9)\},$$

   which has $2 \times 3 = 6$ elements.
2. If $k = 2$ and $A_1 = A_2 = \mathbb{R}$, then $A_1 \times A_2 = \mathbb{R} \times \mathbb{R}$, the set of all ordered pairs of real numbers. Thus $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.
   Similarly, if $A_1, A_2, \ldots, A_k$ are all equal to $\mathbb{R}$, then the product set $A_1 \times A_2 \times \cdots \times A_k = \mathbb{R}^k$.

3. The definition can be extended to the product of infinitely many sets. In particular, the product of infinitely many copies of $\mathbb{R}$ gives the set $\mathbb{R}^\infty$, the set of all ordered infinite tuples of real numbers of the form $(a_1, a_2, a_3, \ldots)$. This is equivalent to the set of all real sequences. We can write these infinite tuples in the form $(a_n)_{n=1}^\infty$, which coincides with our notation for sequences.

This is one reason why we use (a_n) as our notation for sequences, since parentheses imply an ordering whereas braces do not.

4. We often denote an element of a product set $A_1 \times A_2 \times \cdots \times A_k$ as $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and we say that two elements $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_k)$ are equal if $a_i = b_i$ for $i = 1, 2, \ldots, k$.

Exercise 1.1

Write down all the elements of $A_1 \times A_2$ when $A_1 = \{a, b\}$ and $A_2 = \{0, 1, 2, 3\}$.

1.1 Functions and distance

In order to fix notation, we start by discussing briefly points, functions and our usual ideas of distance in $\mathbb{R}^n$, n-dimensional Euclidean space.

Points in $\mathbb{R}^n$ have n coordinates, and are denoted by ordered *n-tuples* $(x_1, x_2, \ldots, x_n)$. As we did with $\mathbb{R}^2$, we use the notation $\mathbf{x}$ for the point $(x_1, x_2, \ldots, x_n)$ and $\mathbf{0}$ for the point $(0, 0, \ldots, 0)$, the **origin** of $\mathbb{R}^n$. Also, if $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$, then the sum and difference of $\mathbf{a}$ and $\mathbf{b}$ are given by

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$

and

$$\mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n)$$

respectively.

Functions between Euclidean spaces will have domain and codomain both Euclidean but possibly of different dimension. Some examples of such functions are:

- $f\colon \mathbb{R}^4 \to \mathbb{R}$ given by $f(x_1, x_2, x_3, x_4) = x_1 + x_2 + x_3 + x_4$, so that, for instance, $f(1, -4, 0, 2) = -1$
- $g\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by $g(x_1, x_2, x_3) = (x_3, x_2)$, so that, for instance, $g(3, 7, 1) = (1, 7)$
- $h\colon \mathbb{R}^5 \to \mathbb{R}^3$ given by $h(x_1, x_2, x_3, x_4, x_5) = (2x_1, x_2^2 + 3x_4, x_3 + x_5)$, so that, for instance, $h(1, -2, 3, -3, 4) = (2, -5, 7)$
- $k\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $k(x_1, x_2) = (\cos 2\pi x_1, \sin \pi x_2, x_1 \sin 2\pi x_2)$, so that, for instance, $k(1, 1.5) = (1, -1, 0)$.

Notice that the domains and codomains of these functions can have different dimensions. In general, we consider functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

In the previous chapter we used the definition of continuity for functions from $\mathbb{R}$ to $\mathbb{R}$ as a basis to define what it meant for functions from $\mathbb{R}^2$ to $\mathbb{R}$ to be continuous. A fundamental observation in this process was that the modulus function acts as the distance function on $\mathbb{R}$; this then led us to consider the distance function $d^{(2)}$ on the plane. In order to define continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, we need to be able to measure the distance between pairs of points in $\mathbb{R}^n$ (and pairs of points in $\mathbb{R}^m$).

Definition 1.2 *Euclidean distance on $\mathbb{R}^n$*

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $\mathbb{R}^n$ is given by the formula

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}.$$

When it is clear from the context which dimension of Euclidean space we are working with, we may simply write $d(\mathbf{a}, \mathbf{b})$ for this distance.

Remarks

1. When $n = 1$ and $n = 2$, we recover the Euclidean distances in $\mathbb{R}$ and $\mathbb{R}^2$, respectively: when $n = 1$, we obtain $d^{(1)}(a, b) = \sqrt{(b-a)^2} = |b - a|$, and when $n = 2$, we obtain the distance between points in the plane.
2. As with $d^{(2)}$, we have $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{a} - \mathbf{b},\ \mathbf{0}) = d^{(n)}(\mathbf{b} - \mathbf{a}, \mathbf{0})$.

For example, in $\mathbb{R}^4$,

$$\begin{aligned} d^{(4)}((1,2,3,4),(2,-1,0,3)) &= \sqrt{(2-1)^2 + (-1-2)^2 + (0-3)^2 + (3-4)^2} \\ &= \sqrt{1+9+9+1} = \sqrt{20} = 2\sqrt{5}. \end{aligned}$$

Exercise 1.2

In $\mathbb{R}^5$, calculate the distance $d^{(5)}((4,2,3,0,-1),(2,-3,-1,0,2))$.

1.2 Properties of the Euclidean distance in $\mathbb{R}^n$

In Subsection 3.4 of Chapter 13 we gave three properties possessed by the Euclidean distance functions on $\mathbb{R}$ and on $\mathbb{R}^2$. In this subsection we verify that the Euclidean distance function $d^{(n)}(\mathbf{a}, \mathbf{b})$ on $\mathbb{R}^n$ has the same properties.

The first property is that distances are never negative and that the distance between two points is zero if, and only if, the two points are the same.

(M1) For each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we have $d^{(n)}(\mathbf{a}, \mathbf{b}) \geq 0$, with $d^{(n)}(\mathbf{a}, \mathbf{b}) = 0$ if, and only if, $\mathbf{a} = \mathbf{b}$.

This is referred to as the *non-negativity property.*

The property that distances are never negative is built into $d^{(n)}$ through the square root. To show that the second part of (M1) holds, suppose first that $\mathbf{a} = \mathbf{b}$. Then $a_i = b_i$ for all values of i, and so each term $(b_i - a_i)^2$ inside the square root of the Euclidean distance function

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}$$

is 0; hence $d^{(n)}(\mathbf{a}, \mathbf{b}) = 0$. Conversely, since each term $(b_i - a_i)^2$ in $d^{(n)}(\mathbf{a}, \mathbf{b})$ is non-negative, when $d^{(n)}(\mathbf{a}, \mathbf{b}) = 0$ each term inside the square root must be equal to 0, and so $a_i = b_i$ for all values of i. Thus $\mathbf{a} = \mathbf{b}$. Therefore, the distance $d^{(n)}(\mathbf{a}, \mathbf{b})$ between two points is 0 *if, and only if,* $\mathbf{a} = \mathbf{b}$.

The next property is that the distance from $\mathbf{a}$ to $\mathbf{b}$ is the same as the distance from $\mathbf{b}$ to $\mathbf{a}$.

(M2) For each $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we have $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{a})$.

This is referred to as the *symmetry property.*

Exercise 1.3

Verify property (M2) by examining $d^{(n)}(\mathbf{a}, \mathbf{b})$ and $d^{(n)}(\mathbf{b}, \mathbf{a})$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.

The symmetry property of $d^{(n)}(\mathbf{a}, \mathbf{b})$ allows us to use the following two forms interchangeably:

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2},$$

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}.$$

The third property is the **Triangle Inequality**.

(M3) For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$, we have $d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$.

There are many different proofs of the Triangle Inequality. We defer a proof of it to Subsection 1.5, where we will give one that is purely algebraic.

We summarise these results in the following proposition.

Proposition 1.3 *Properties of Euclidean distance on $\mathbb{R}^n$*

The Euclidean distance function $d^{(n)}$ on $\mathbb{R}^n$ has the following properties.

For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$:

non-negativity (M1) $d^{(n)}(\mathbf{a}, \mathbf{b}) \geq 0$, with equality holding if, and only if, $\mathbf{a} = \mathbf{b}$

symmetry (M2) $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{a})$

Triangle Inequality (M3) $d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$.

In Section 2, we will use these three properties as the definition for a general distance function, so you need to be familiar with them. In particular, be sure that you understand the meaning of the Triangle Inequality, as it is of basic importance.

The Triangle Inequality for $d^{(n)}$, like that for distances in $\mathbb{R}$, has a 'reversed' form.

Proposition 1.4 *Reverse Triangle Inequality for $\mathbb{R}^n$*

The Euclidean distance function $d^{(n)}$ on $\mathbb{R}^n$ has the following property.

(M3a) If $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$, then $d^{(n)}(\mathbf{b}, \mathbf{c}) \geq \left| d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \right|$.

Proof From (M3) we obtain, for each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$,

$$d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c}).$$

Subtracting $d^{(n)}(\mathbf{a}, \mathbf{b})$ from each side, we obtain

$$d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \leq d^{(n)}(\mathbf{b}, \mathbf{c}). \tag{1}$$

Since $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are arbitrary points in $\mathbb{R}^n$, we can freely interchange their roles in the Triangle Inequality. Thus, interchanging the roles of $\mathbf{b}$ and $\mathbf{c}$ in equation (1), we get

$$d^{(n)}(\mathbf{a}, \mathbf{b}) - d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{c}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{c}), \text{ by (M2)}.$$

This can be rearranged to give

$$-d^{(n)}(\mathbf{b}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}). \tag{2}$$

Combining inequalities (1) and (2), we obtain

$$-d^{(n)}(\mathbf{b}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \leq d^{(n)}(\mathbf{b}, \mathbf{c}),$$

which is equivalent to

$$\left| d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \right| \leq d^{(n)}(\mathbf{b}, \mathbf{c}),$$

as required. ■

We finish this subsection by examining the domain and codomain of the distance function determined by the Euclidean distance for $\mathbb{R}^n$. The Euclidean distance formula

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}$$

defines a function of two variables $\mathbf{a}$ and $\mathbf{b}$, each from the set $\mathbb{R}^n$, which produce a (non-negative) real number, the distance.

The domain of the distance function $d^{(1)}$ is simply the set of all ordered pairs (a, b) where $a \in \mathbb{R}$ and $b \in \mathbb{R}$ – that is, the domain is the product set $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. However, each of the arguments $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ of the distance function $d^{(2)}$ is itself an ordered pair of real numbers from $\mathbb{R}^2$. Thus the domain of the distance function $d^{(2)}$ is the set of all ordered pairs of elements of $\mathbb{R}^2$, i.e. the product set $\mathbb{R}^2 \times \mathbb{R}^2$. In the general case, the domain of the distance function $d^{(n)}$ is the product set $\mathbb{R}^n \times \mathbb{R}^n$.

For each n, we take the codomain of $d^{(n)}$ to be $\mathbb{R}$.

We could take the codomain to be any set that contains $[0, \infty)$.

Hence $d^{(n)}$ is a function

$$d^{(n)} \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

from the set of all ordered pairs of ordered n-tuples of real numbers to the set of real numbers.

1.3 Sequences in $\mathbb{R}^n$

In the previous chapter we studied real sequences and sequences of points in $\mathbb{R}^2$. We now extend these ideas to include sequences in $\mathbb{R}^n$. Following Chapter 13, we will use convergence of sequences (in $\mathbb{R}^n$ and $\mathbb{R}^m$) to define the continuity of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ (in the next subsection).

The definitions of sequences and convergence in $\mathbb{R}^n$ are very similar to those in the plane: we simply change the set from $\mathbb{R}^2$ to $\mathbb{R}^n$ and replace the Euclidean distance $d^{(2)}$ with the n-dimensional version $d^{(n)}$.

Definition 1.5 *Sequence in $\mathbb{R}^n$*

A **sequence** in $\mathbb{R}^n$ is an unending ordered list of points in $\mathbb{R}^n$:

$$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots.$$

The point $\mathbf{a}_k = (a_{1,k}, a_{2,k}, \ldots, a_{n,k})$ is the ***k*th term** of the sequence, and the whole sequence is denoted by $(\mathbf{a}_k)$, $(\mathbf{a}_k)_{k=1}^{\infty}$ or $(\mathbf{a}_k)_{k \in \mathbb{N}}$.

Here we use the letter k for a general term in the sequence as the letter n has already been used to denote the dimension of $\mathbb{R}^n$.

Definition 1.6 *Convergence in $\mathbb{R}^n$*

A sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$ **converges** to $\boldsymbol{\ell} \in \mathbb{R}^n$ if $\left(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})\right)$ is a (real) null sequence.

We say that $\boldsymbol{\ell}$ is the **limit** of the sequence $(\mathbf{a}_k)$ and write $\mathbf{a}_k \to \boldsymbol{\ell}$ as $k \to \infty$, $\lim_{k\to\infty} \mathbf{a}_k = \boldsymbol{\ell}$ or simply $\mathbf{a}_k \to \boldsymbol{\ell}$.

A sequence in $\mathbb{R}^n$ that does not converge to any point in $\mathbb{R}^n$ is said to be **divergent**.

Remarks

1. A sequence in $\mathbb{R}^n$ that converges to $\mathbf{0}$ is called a **null sequence**.
2. Note that $\left(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})\right)_{k\in\mathbb{N}}$ is a sequence of real numbers. Hence, testing whether a given sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$ converges to $\boldsymbol{\ell} \in \mathbb{R}^n$ reduces to testing whether $\left(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})\right)$ is a *real* null sequence.
3. Definition 1.6 is equivalent to the following:

   A sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$ converges to $\boldsymbol{\ell}$ if, for each $\varepsilon > 0$, there is a $K \in \mathbb{N}$ such that

   $$d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell}) < \varepsilon \quad \text{whenever } k > K.$$

There is a very important fact about convergent sequences in $\mathbb{R}^n$, which we will prove now before we consider some examples: limits of convergent sequences are unique.

Lemma 1.7 *Uniqueness of limits in $\mathbb{R}^n$*

Let $(\mathbf{a}_k)$ be a convergent sequence in $\mathbb{R}^n$. If $\mathbf{a}_k \to \boldsymbol{r}$ and $\mathbf{a}_k \to \boldsymbol{s}$ as $k \to \infty$, then $\boldsymbol{r} = \boldsymbol{s}$.

Proof We let d denote the Euclidean distance on $\mathbb{R}^n$ and show that $d(\boldsymbol{r}, \boldsymbol{s}) = 0$. Property (M1) for d then implies that $\boldsymbol{r} = \boldsymbol{s}$, which proves the lemma.

The definition of convergence for sequences in $\mathbb{R}^n$ tells us that $(d(\mathbf{a}_k, \boldsymbol{r}))$ and $(d(\mathbf{a}_k, \boldsymbol{s}))$ are both *real* null sequences. Hence we can use property (M3) for d, the Triangle Inequality, to estimate that for each $k \in \mathbb{N}$,

$$\begin{aligned} d(\boldsymbol{r}, \boldsymbol{s}) &\leq d(\boldsymbol{r}, \mathbf{a}_k) + d(\mathbf{a}_k, \boldsymbol{s}) \\ &= d(\mathbf{a}_k, \boldsymbol{r}) + d(\mathbf{a}_k, \boldsymbol{s}), \text{ by property (M2) for } d. \end{aligned}$$

However, both $d(\mathbf{a}_k, \boldsymbol{r}) \to 0$ and $d(\mathbf{a}_k, \boldsymbol{s}) \to 0$ as $k \to \infty$ and so

$$d(\boldsymbol{r}, \boldsymbol{s}) \leq 0.$$

But it is certainly the case (by property (M1) for d) that $d(\boldsymbol{r}, \boldsymbol{s}) \geq 0$ and so $d(\boldsymbol{r}, \boldsymbol{s}) = 0$, as required. ■

Notice that the proof of this lemma only uses the fact that the Euclidean distance on $\mathbb{R}^n$ has properties (M1)–(M3).

Worked Exercise 1.8

Prove that the sequence $(\mathbf{b}_k)$ in $\mathbb{R}^3$ given by $\mathbf{b}_k = (\frac{k+1}{k}, \frac{(-1)^k}{k}, 3)$ for each $k \in \mathbb{N}$ converges, and find its limit.

Solution

Since $\frac{k+1}{k} \to 1$ and $\frac{(-1)^k}{k} \to 0$ as $k \to \infty$, the terms of $(\mathbf{b}_k)$ appear to get near to $(1, 0, 3)$ as k gets large. So letting d denote the Euclidean distance on $\mathbb{R}^3$, then for $k \in \mathbb{N}$, we calculate

$$\begin{aligned} d(\mathbf{b}_k, (1,0,3)) &= \sqrt{\left(1 - \tfrac{k+1}{k}\right)^2 + \left(0 - \tfrac{(-1)^k}{k}\right)^2 + (3-3)^2} \\ &= \sqrt{\tfrac{1}{k^2} + \tfrac{1}{k^2}} = \tfrac{\sqrt{2}}{k}. \end{aligned}$$

Since $\left(\frac{\sqrt{2}}{k}\right)$ is a multiple of a basic null sequence, we conclude that $(d(\mathbf{b}_k, (1,0,3)))$ is a real null sequence and so $(\mathbf{b}_k)$ converges to $(1, 0, 3)$ as $k \to \infty$ in $\mathbb{R}^3$.

Exercise 1.4

(a) Let the sequence $(\mathbf{a}_k)$ in $\mathbb{R}^4$ be defined by

$$\mathbf{a}_k = \left(\frac{-1}{(-3)^k}, \frac{3^k + 5\cos k\pi}{3^k}, \sin k\pi, -1 - \frac{2}{3^k}\right).$$

Show that $(\mathbf{a}_k)$ converges, and find its limit.

(b) Prove that any constant sequence in $\mathbb{R}^n$ converges.

If $(\mathbf{a}_k)$ is a sequence in $\mathbb{R}^n$, then the n sequences $(a_{j,k})$ for $j = 1, 2, \ldots, n$ can be thought of as individual sequences of real numbers in their own right. We refer to them as **coordinate sequences**. The next proposition makes clear the connection between the convergence of a sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$ and the convergence of these n individual coordinate sequences.

Proposition 1.9

Let $(\mathbf{a}_k)$ be a sequence in $\mathbb{R}^n$, with $\mathbf{a}_k = (a_{1,k}, \ldots, a_{n,k})$, and let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n) \in \mathbb{R}^n$. Then

$$\mathbf{a}_k \to \boldsymbol{\ell} \quad \text{as } k \to \infty$$

if, and only if,

$$a_{j,k} \to \ell_j \quad \text{as } k \to \infty, \text{ for each } j = 1, 2, \ldots, n.$$

Proposition 1.9 extends the result of Theorem 3.7 from Chapter 13.

Proposition 1.9 is useful because it tells us that to test the convergence of a given sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$, it is enough to test the convergence of the real coordinate sequences $(a_{j,k})$ for each j. We give some applications of this proposition before we ask you to prove it in Exercise 1.7.

Worked Exercise 1.10

Use Proposition 1.9 to prove the Sum Rule for convergent sequences in $\mathbb{R}^n$: if $\mathbf{a}_k \to \boldsymbol{\ell}$ and $\mathbf{b}_k \to \mathbf{m}$ as $k \to \infty$, then $\mathbf{a}_k + \mathbf{b}_k \to \boldsymbol{\ell} + \mathbf{m}$ as $k \to \infty$.

Solution

Let $(\mathbf{a}_k)$ and $(\mathbf{b}_k)$ be sequences in $\mathbb{R}^n$ such that $\mathbf{a}_k \to \boldsymbol{\ell}$ and $\mathbf{b}_k \to \mathbf{m}$. We need to show that $(\mathbf{a}_k + \mathbf{b}_k)$ converges to $\boldsymbol{\ell} + \mathbf{m}$.

Proposition 1.9 gives that

$$a_{j,k} \to \ell_j \quad \text{and} \quad b_{j,k} \to m_j \quad \text{as } k \to \infty, \text{ for } j = 1, 2, \ldots, n.$$

By applying the Sum Rule for convergent real sequences, we see that

$$a_{j,k} + b_{j,k} \to \ell_j + m_j \quad \text{as } k \to \infty, \text{ for } j = 1, 2, \ldots, n.$$

Using Proposition 1.9 again now shows that $\mathbf{a}_k + \mathbf{b}_k \to \boldsymbol{\ell} + \mathbf{m}$ as $k \to \infty$.

Exercise 1.5

Prove the Multiple Rule for a convergent sequence $\mathbf{a}_k \to \boldsymbol{\ell}$ in $\mathbb{R}^n$. That is, show that $\lambda \mathbf{a}_k \to \lambda \boldsymbol{\ell}$ for any $\lambda \in \mathbb{R}$.

For $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, we write $\lambda \mathbf{x} = (\lambda x_1, \ldots, \lambda x_n)$.

Exercise 1.6

Use Proposition 1.9 to show that the sequence $(\mathbf{a}_k)$ in $\mathbb{R}^3$ given by

$$\mathbf{a}_k = \left(\frac{k^3 + 3^k}{k!}, 1 - \frac{1}{\sqrt{k}}, 2^{-k}(5 - 2^k) \right)$$

converges, and find its limit.

Exercise 1.7

Adapt the proof of Theorem 3.7 from Chapter 13 to prove Proposition 1.9.

For convenience, we summarise the combination rules that we have derived in these exercises.

Combination rules for convergent sequences in $\mathbb{R}^n$

If $\mathbf{a}_k \to \boldsymbol{\ell}$ and $\mathbf{b}_k \to \mathbf{m}$ as $k \to \infty$, then:

Sum Rule $\mathbf{a}_k + \mathbf{b}_k \to \boldsymbol{\ell} + \mathbf{m}$ as $k \to \infty$

Multiple Rule $\lambda \mathbf{a}_k \to \lambda \boldsymbol{\ell}$ as $k \to \infty$, for any $\lambda \in \mathbb{R}$.

We do not give a Product Rule for sequences in $\mathbb{R}^n$ because we have not defined what it means to 'multiply' points of $\mathbb{R}^n$ together.

Proposition 1.9 also provides a very convenient method for showing that a sequence in $\mathbb{R}^n$ diverges. It tells us that a sequence $(\mathbf{a}_k) = ((a_{1,k}, \ldots, a_{n,k}))$ is divergent if *any one* of the coordinate sequences $(a_{j,k})$ is divergent. For example, the sequence in $\mathbb{R}^3$ given by $\mathbf{a}_k = (\frac{1}{k}, \frac{2k^2+1}{k^2}, (-1)^k)$ diverges because the third coordinate sequence $(a_{3,k}) = ((-1)^k)$ does not converge to any real number.

The sequence $((-1)^k)$ was shown to be divergent in Worked Exercise 1.26 of Chapter 13.

Exercise 1.8

Let $(\mathbf{b}_k)$ be the sequence in $\mathbb{R}^4$ given by $\mathbf{b}_k = \left(\dfrac{k!}{k^k}, k, \dfrac{2}{\sqrt{k}}, k^{1/\log k}\right)$.

Determine whether $(\mathbf{b}_k)$ converges, justifying your answer.

1.4 Continuity for functions between Euclidean spaces

We are now ready to extend the definition of continuity to functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. Given all our work up to this point, it is not too difficult to see what we have to do to define continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$: we just have to work with sequences in $\mathbb{R}^n$ in the domain and sequences in $\mathbb{R}^m$ in the codomain.

The resulting definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ mimics that for functions from $\mathbb{R}^2$ to $\mathbb{R}$, which was given in Chapter 13.

See Chapter 13, Definition 3.8.

Definition 1.11 *Continuity from $\mathbb{R}^n$ to $\mathbb{R}^m$*

Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be a function.

Then f is **continuous** at $\mathbf{a} \in A$ if:

whenever $(\mathbf{x}_k)$ is a sequence in A for which $\mathbf{x}_k \to \mathbf{a}$ as $k \to \infty$, then the sequence $(f(\mathbf{x}_k))$ converges to $f(\mathbf{a})$.

If f does not satisfy this condition at $\mathbf{a} \in A$ – that is, there is a sequence $(\mathbf{x}_k)$ in A for which $\mathbf{x}_k \to \mathbf{a}$ as $k \to \infty$ but $f(\mathbf{x}_k)$ does not converge to $f(\mathbf{a})$ – then we say that f is **discontinuous** at $\mathbf{a}$.

Note that the sequence $(f(\mathbf{x}_k))$ in the above definition is a sequence in $\mathbb{R}^m$.

Definition 1.12 *Continuity on a set in $\mathbb{R}^n$*

Let $S \subseteq A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

It is worth reflecting briefly on how we have arrived at this definition of continuity. We first examined the notion of distance, extending the distance function $d^{(1)}(a, b) = |b - a|$ on $\mathbb{R}$ to obtain the Euclidean distance function $d^{(n)}(\mathbf{a}, \mathbf{b})$ on $\mathbb{R}^n$. Next, we used this idea of distance to provide a definition of the convergence of sequences in higher dimensions. Finally, we have used convergent sequences to say what it means for a function from $\mathbb{R}^n$ to $\mathbb{R}^m$ to be continuous. So we see that it is our ideas of distance that underpin the concept of continuity.

We now consider some examples of continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$, for various values of n and m.

Worked Exercise 1.13

Prove that the function $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$f(x_1, x_2, x_3) = (x_1 + x_3, x_2)$$

is continuous on $\mathbb{R}^3$.

Solution

Let $\mathbf{a}$ be a (general) point of $\mathbb{R}^3$ and let $(\mathbf{x}_k)$ be a sequence in $\mathbb{R}^3$ such that $\mathbf{x}_k \to \mathbf{a}$ as $k \to \infty$.

Writing $\mathbf{x}_k = (x_{1,k}, x_{2,k}, x_{3,k})$ and $\mathbf{a} = (a_1, a_2, a_3)$, Proposition 1.9 tells us that

$$x_{1,k} \to a_1, \quad x_{2,k} \to a_2 \quad \text{and} \quad x_{3,k} \to a_3 \quad \text{as } k \to \infty. \tag{3}$$

We need to show that the sequence $(f(\mathbf{x}_k))$ converges to $f(\mathbf{a})$ in $\mathbb{R}^2$, where

$$f(\mathbf{x}_k) = f(x_{1,k}, x_{2,k}, x_{3,k}) = (x_{1,k} + x_{3,k}, x_{2,k}).$$

Applying the Sum Rule (for real sequences) to equation (3) gives that $x_{1,k} + x_{3,k} \to a_1 + a_3$ as $k \to \infty$. Therefore, using equation (3) and Proposition 1.9 again, it follows that

$$f(\mathbf{x}_k) = (x_{1,k} + x_{3,k}, x_{2,k}) \to (a_1 + a_3, a_2) = f(\mathbf{a}) \quad \text{as } k \to \infty.$$

We conclude that f is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^3$, f is continuous on $\mathbb{R}^3$.

Worked Exercise 1.14

Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f(\mathbf{x}) = (\text{int}(2x_1), 3x_2),$$

where, as usual, $\text{int}(a)$ denotes the integer part of a (the largest integer less than or equal to a). Prove that f is a discontinuous function.

Solution

We need to find a point at which the function f is discontinuous. That is, we must find $\mathbf{a} \in \mathbb{R}^2$ and a sequence $(\mathbf{x}_k)$ of points in the plane for which $\mathbf{x}_k \to \mathbf{a}$ and yet $f(\mathbf{x}_k) \not\to f(\mathbf{a})$ as $k \to \infty$.

It turns out that there are many possible choices for $\mathbf{a}$. Consider the point $\mathbf{a} = (\frac{1}{2}, 0)$. Then $f(\mathbf{a}) = (\text{int}(1), 0) = (1, 0)$ but if $\mathbf{x}$ is a point with x_1-coordinate just a little less than $\frac{1}{2}$, then $\text{int}(2x_1) = 0$. Hence if we take a sequence of points in the plane $(\mathbf{x}_k)$ that converge to $\mathbf{a}$ from the 'left' (Figure 1.1), then $f(\mathbf{x}_k)$ will have first coordinate zero for large enough k and so cannot converge to $f(\mathbf{a})$. For example, take $\mathbf{x}_k = \left(\frac{1}{2} - \frac{1}{2k}, 0\right)$. Then certainly $\mathbf{x}_k \to (\frac{1}{2}, 0) = \mathbf{a}$ as $k \to \infty$ but for each $k \in \mathbb{N}$, $f(\mathbf{x}_k) = (\text{int}(1 - \frac{1}{k}), 0) = (0, 0) \not\to (1, 0) = f(\mathbf{a})$. Hence f is discontinuous at $\mathbf{a}$.

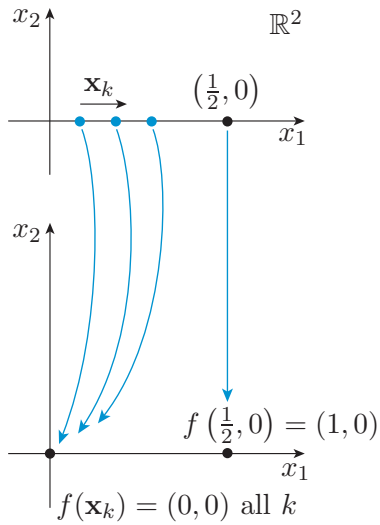


Figure 1.1 Sequence of points converging to $\mathbf{a}$ from the left

In fact f is discontinuous at any point of the form $(\frac{m}{2}, x_2)$ for $m \in \mathbb{Z}$ and $x_2 \in \mathbb{R}$.

Exercise 1.9

Let $\mathbf{c} \in \mathbb{R}^m$ be fixed. Show that the constant function $f : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$f(\mathbf{x}) = \mathbf{c}$$

is continuous on $\mathbb{R}^n$.

In the next worked exercise we apply the definition of continuity to the distance function $d^{(n)}(\mathbf{x}, \mathbf{u})$ that maps a point in $\mathbb{R}^n$ to its distance from a fixed point $\mathbf{u}$; it is thus a function from $\mathbb{R}^n$ to $\mathbb{R}$.

Worked Exercise 1.15

Let $\mathbf{u}$ be a fixed point in $\mathbb{R}^n$. Prove that the function $g : \mathbb{R}^n \to \mathbb{R}$ given by

$$g(\mathbf{x}) = d^{(n)}(\mathbf{x}, \mathbf{u})$$

is continuous on $\mathbb{R}^n$. (The graph of this function is shown in Figure 1.2.)

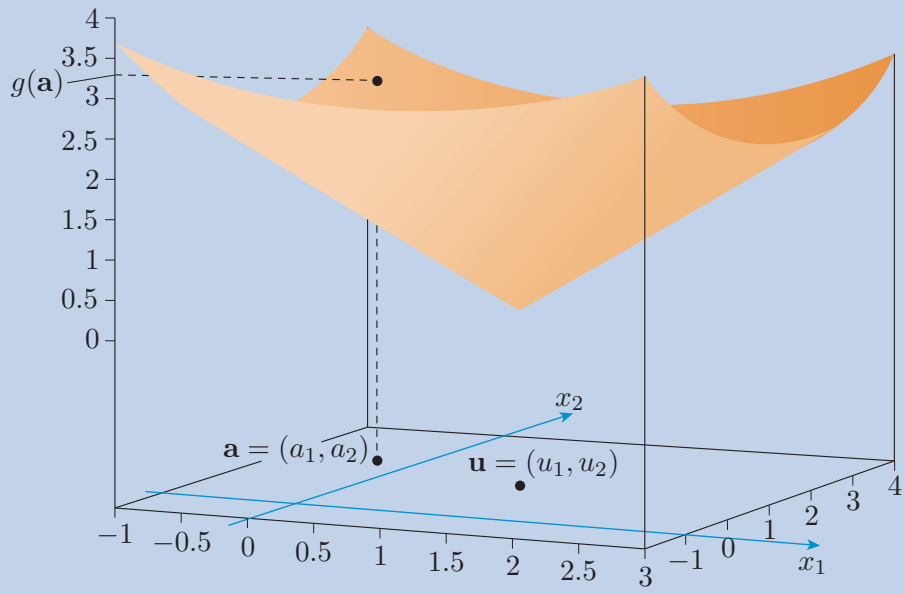


Figure 1.2 Graph of function for Worked Exercise 1.15

Solution

Let $\mathbf{a}$ be a general point of $\mathbb{R}^n$ and let $(\mathbf{x}_k)$ be a sequence in $\mathbb{R}^n$ such that $\mathbf{x}_k \to \mathbf{a}$. We need to show that $(d^{(1)}(g(\mathbf{x}_k), g(\mathbf{a})))$ is a null sequence in $\mathbb{R}$.

We find an upper bound for $d^{(1)}(g(\mathbf{x}_k), g(\mathbf{a})) = |g(\mathbf{x}_k) - g(\mathbf{a})|$ in terms of $d^{(n)}(\mathbf{x}_k, \mathbf{a})$. By the definition of g,

$$|g(\mathbf{x}_k) - g(\mathbf{a})| = |d^{(n)}(\mathbf{x}_k, \mathbf{u}) - d^{(n)}(\mathbf{a}, \mathbf{u})|.$$

From the Reverse Triangle Inequality, Proposition 1.4, we see that

$$|\, d^{(n)}(\mathbf{x}_k, \mathbf{u}) - d^{(n)}(\mathbf{a}, \mathbf{u})| = |\, d^{(n)}(\mathbf{u}, \mathbf{x}_k) - d^{(n)}(\mathbf{u}, \mathbf{a})| \leq d^{(n)}(\mathbf{x}_k, \mathbf{a}).$$

So, $0 \leq |g(\mathbf{x}_k) - g(\mathbf{a})| \leq d^{(n)}(\mathbf{x}_k, \mathbf{a})$.

But $\mathbf{x}_k \to \mathbf{a}$, and so $\big(d^{(n)}(\mathbf{x}_k, \mathbf{a})\big)$ is a (real) null sequence. Hence, the Squeeze Rule implies that $g(\mathbf{x}_k) \to g(\mathbf{a})$.

Thus g is continuous at $\mathbf{a}$. Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^n$, g is continuous on $\mathbb{R}^n$.

In this solution we need to keep a clear head, since the distance function is used in the domain and also as the function in the codomain.

As with the functions discussed in the previous chapter, there are combination rules for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ and these often make it easier to show that a given function is continuous.

Combination rules for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let $A \subseteq \mathbb{R}^n$, and let $f\colon A \to \mathbb{R}^m$ and $g\colon A \to \mathbb{R}^m$ be continuous on A.

Then the following functions are continuous on A:

Sum Rule $f + g\colon A \to \mathbb{R}^m$, defined by $(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$

Multiple Rule $\lambda f\colon A \to \mathbb{R}^m$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$.

These rules can be quickly deduced from the combination rules for convergent sequences in higher dimensions.

There is also a Restriction Rule for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ that allows us to restrict the domain of a continuous function to a smaller set and still have a continuous function.

Proposition 1.16 *Restriction Rule*

Let $A \subseteq \mathbb{R}^n$, let $f\colon A \to \mathbb{R}^m$ be continuous on A and let $B \subseteq A$. Then the restricted function $f|_B\colon B \to \mathbb{R}^m$ defined by $f|_B(\mathbf{x}) = f(\mathbf{x})$ for $\mathbf{x} \in B$ is continuous on B.

We can also generalise the Composition Rule. If the first function is from $\mathbb{R}^n$ to $\mathbb{R}^m$, then the second function must have its domain in $\mathbb{R}^m$, but may have any dimension for its codomain. Again we state it without proof.

Proposition 1.17 *Composition Rule*

Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}^m$ and let $g\colon B \to \mathbb{R}^k$ be continuous on B. Then the composed function $g \circ f\colon A \to \mathbb{R}^k$ is continuous on A.

Exercise 1.10

Without using sequences, prove that the function $g\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$g(\mathbf{x}) = d^{(n)}(\mathbf{x}, \mathbf{0})^2 \quad \left(= d^{(n)}(\mathbf{x}, \mathbf{0}) \times d^{(n)}(\mathbf{x}, \mathbf{0})\right)$$

This function maps $\mathbf{x}$ to the square of its distance from $\mathbf{0}$.

is continuous on $\mathbb{R}^n$.

The Composition Rule for continuous functions is particularly useful when used in combination with projection functions from $\mathbb{R}^n$ to $\mathbb{R}$, which are similar to the projection functions from $\mathbb{R}^2$ to $\mathbb{R}$ introduced in Chapter 13.

See Chapter 13, Definition 3.10.

Definition 1.18 *Projection functions on $\mathbb{R}^n$*

The n functions $p_j\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$p_j(x_1, x_2, \ldots, x_n) = x_j, \quad \text{for } j = 1, 2, \ldots, n,$$

are known as **projection functions**.

Each of these projection functions maps a point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ onto one of its coordinates. Thus $p_1(\mathbf{x}) = x_1$, $p_2(\mathbf{x}) = x_2$, and so on.

Exercise 1.11

Show that the projection functions are continuous on $\mathbb{R}^n$.

Note that the effect of composing each separate p_j with a function f is to produce the individual component function values of f in the codomain. For example, composing p_1, p_2 and p_3 with the function

$$h\colon \mathbb{R}^5 \to \mathbb{R}^3, \quad h(x_1, x_2, x_3, x_4, x_5) = (2x_1, x_2^2 + 3x_4, x_3 + x_5),$$

gives

$$(p_1 \circ h)(x_1, x_2, x_3, x_4, x_5) = 2x_1,$$
$$(p_2 \circ h)(x_1, x_2, x_3, x_4, x_5) = x_2^2 + 3x_4,$$
$$(p_3 \circ h)(x_1, x_2, x_3, x_4, x_5) = x_3 + x_5.$$

The Composition Rule and the projection functions $p_j : \mathbb{R}^m \to \mathbb{R}$ give us a useful technique for proving that a function f from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous. The main idea is that, instead of having to prove directly that f is continuous, we form the composition of each of the projection functions p_j with the function f, and then prove that each of these functions $p_j \circ f$ is continuous. Since the functions $p_j \circ f$ are functions from $\mathbb{R}^m$ to $\mathbb{R}$, this is usually simpler than proving directly that f is continuous. The following result enables us to use projection functions in this way.

Proposition 1.19

Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$.

The function f is continuous at $\mathbf{a} \in A$ if, and only if, for each $j = 1, 2, \ldots, m$, the function $p_j \circ f : A \to \mathbb{R}$ is continuous at $\mathbf{a}$, where $p_j : \mathbb{R}^m \to \mathbb{R}$ is the jth projection function.

Proof The proof is in two parts. In each part we deal with the m functions $p_1 \circ f, p_2 \circ f, \ldots, p_m \circ f$ simultaneously.

First, we prove that if f is continuous at $\mathbf{a}$ then the functions $p_j \circ f$ are also continuous at $\mathbf{a}$. This follows directly from the Composition Rule for continuous functions. Since f is continuous at $\mathbf{a}$ and each of the p_j are continuous on $\mathbb{R}^m$, the functions $p_j \circ f$ are also continuous at $\mathbf{a}$, for $j = 1, 2, \ldots, m$.

Conversely, we now prove that if the functions $p_j \circ f$ are continuous at $\mathbf{a}$ then f is also continuous at $\mathbf{a}$. To do this, let $(\mathbf{x}_k)$ be a sequence in A that converges to $\mathbf{a}$. Then, since the functions $p_j \circ f$ are continuous at $\mathbf{a}$, we have that

This is the useful part of Proposition 1.19.

$$(p_j \circ f)(\mathbf{x}_k) \to (p_j \circ f)(\mathbf{a}) \text{ as } k \to \infty \quad \text{for each } j = 1, 2, \ldots, m.$$

The value of f at any point $\mathbf{x} \in A$ can be expressed as

$$f(\mathbf{x}) = ((p_1 \circ f)(\mathbf{x}), \ldots, (p_m \circ f)(\mathbf{x})).$$

Using Proposition 1.9, it follows that

$$\begin{aligned} f(\mathbf{x}_k) &= ((p_1 \circ f)(\mathbf{x}_k), \ldots, (p_m \circ f)(\mathbf{x}_k)) \\ &\to ((p_1 \circ f)(\mathbf{a}), \ldots, (p_m \circ f)(\mathbf{a})) = f(\mathbf{a}) \text{ as } k \to \infty. \end{aligned}$$

Thus f is continuous at $\mathbf{a}$. ■

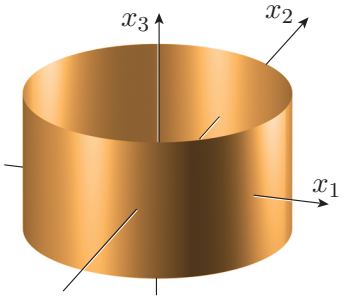


Figure 1.3 The image set $f(A)$ in Worked Exercise 1.20

See Worked Exercise 3.14 in Chapter 13.

Worked Exercise 1.20

Let $A = [0,1] \times [-1,1] = \{(x_1, x_2) : 0 \leq x_1 \leq 1 \text{ and } -1 \leq x_2 \leq 1\}$ and define $f : A \to \mathbb{R}^3$ by

$$f(x_1, x_2) = (\cos 2\pi x_1, \sin 2\pi x_1, x_2).$$

Prove that f is continuous on A. (Note that the image set $f(A)$ is a cylinder, as shown in Figure 1.3.)

Solution

We consider the composition of f with the three projection functions p_1, p_2 and p_3 from $\mathbb{R}^3$ to $\mathbb{R}$, given by

$$(p_1 \circ f)(x_1, x_2) = \cos 2\pi x_1,$$
$$(p_2 \circ f)(x_1, x_2) = \sin 2\pi x_1,$$
$$(p_3 \circ f)(x_1, x_2) = x_2.$$

If we can show that each $p_j \circ f$ is continuous on A, then it follows from Proposition 1.19 that f is continuous on A.

To show that $(p_1 \circ f)(x_1, x_2) = \cos 2\pi x_1$ is continuous on A, we show that the constituent parts are continuous and then combine them using the combination rules.

In the previous chapter, we proved the continuity of the function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $g(\mathbf{x}) = u_1 x_1 + u_2 x_2$, where u_1 and u_2 are fixed. It follows, on taking $u_1 = 2\pi$ and $u_2 = 0$, that the function given by $(x_1, x_2) \mapsto 2\pi x_1$ is continuous.

We are also restricting the functions to the set A, and so we need to apply the Restriction Rule for functions from $\mathbb{R}^2$ to $\mathbb{R}$. Then, since $x \mapsto \cos x$ is a basic continuous function from $\mathbb{R}$ to $\mathbb{R}$, we can apply the Composition Rule for continuous functions. This shows that $(x_1, x_2) \mapsto \cos 2\pi x_1$ is continuous.

The proof of continuity for $p_2 \circ f$ is similar, except that we use the basic continuous function $x \mapsto \sin x$.

Finally, we note that $p_3 \circ f$ is a function of the form $g(\mathbf{x}) = u_1 x_1 + u_2 x_2$, as considered above, with $\mathbf{u} = (0, 1)$ and restricted to the set A. Hence, by the Restriction Rule, it is continuous. (Alternatively we could simply note that $p_3 \circ f$ is the projection function $p_2 : \mathbb{R}^2 \to \mathbb{R}$ and hence is continuous.)

Since all three compositions of f with a projection function are continuous on A, we conclude from Proposition 1.19 that f is continuous on A.

Exercise 1.12

Let $\mathbf{v} = (v_1, v_2, v_3)$ be a fixed point in $\mathbb{R}^3$. Define $f : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$f(x_1, x_2, x_3) = (v_3x_2 - v_2x_3, v_1x_3 - v_3x_1, v_2x_1 - v_1x_2).$$

Prove that f is continuous on $\mathbb{R}^3$.

1.5 Proof of the Triangle Inequality

This subsection is not assessed.

In this subsection we give a proof of the Triangle Inequality

$$d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$$

using only algebraic methods. The proof makes use of another inequality, the *Cauchy–Schwarz Inequality*, which has important applications in many areas of mathematics. We first state and prove this inequality.

Theorem 1.21 *Cauchy–Schwarz Inequality*

Let $(r_1, r_2, \ldots, r_n)$ and $(s_1, s_2, \ldots, s_n)$ be points in $\mathbb{R}^n$. Then

$$\left(\sum_{j=1}^{n} r_j s_j\right)^2 \leq \sum_{j=1}^{n} r_j^2 \sum_{j=1}^{n} s_j^2, \tag{4}$$

with equality if, and only if, there is $\lambda \in \mathbb{R}$ for which $s_j = \lambda r_j$ for each j.

A useful technique for understanding a general statement is to specialise to small particular values.

When $n = 1$, the right-hand side of the Cauchy–Schwarz Inequality is $(r_1^2)(s_1^2)$ and the left-hand side is $(r_1 s_1)^2$. So taking the left-hand side from the right-hand side gives

$$(r_1^2)(s_1^2) - (r_1 s_1)^2 = 0,$$

and so inequality (4) certainly holds for $n = 1$.

When $n = 2$, the right-hand side becomes

$$(r_1^2 + r_2^2)(s_1^2 + s_2^2)$$

and the left-hand side is

$$(r_1 s_1 + r_2 s_2)^2.$$

Thus subtracting the left-hand side from the right-hand side gives

$$(r_1^2 + r_2^2)(s_1^2 + s_2^2) - (r_1 s_1 + r_2 s_2)^2$$

and, multiplying out, we obtain

$$\begin{aligned}&(r_1^2s_1^2 + r_1^2s_2^2 + r_2^2s_1^2 + r_2^2s_2^2) - (r_1^2s_1^2 + 2r_1r_2s_1s_2 + r_2^2s_2^2)\\&= r_1^2s_2^2 + r_2^2s_1^2 - 2r_1r_2s_1s_2\\&= (r_1s_2 - r_2s_1)^2 \geq 0.\end{aligned}$$

So (4) holds (with equality if, and only if, there is $\lambda \in \mathbb{R}$ for which $s_1 = \lambda r_1$ and $s_2 = \lambda r_2$).

We could give a proof of the Cauchy–Schwarz Inequality by developing these ideas for a general value of n. However, we give a different proof, which involves a strategy based on some properties of quadratic equations with which you may well be familiar.

Let $A, B, C \in \mathbb{R}$, where $A \neq 0$, and suppose that

$$Ax^2 + Bx + C \geq 0, \quad \text{for each } x \in \mathbb{R}.$$

Then the quadratic equation $Ax^2 + Bx + C = 0$ has at most one real root. For, if the equation had two distinct real roots, then the graph of $f(x) = Ax^2 + Bx + C$ would cut the x-axis in two places and there would be a range of x for which $f(x) < 0$, i.e. $Ax^2 + Bx + C < 0$.

The roots of the quadratic equation are given by

$$x = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

When there is at most one real root, it must be the case that $B^2 - 4AC \leq 0$, so that $B^2 \leq 4AC$. In particular, there is exactly one real root in the case that $B^2 = 4AC$; otherwise there are none. This is the idea we use.

Proof of the Cauchy–Schwarz Inequality If $r_j = 0$ for $j = 1, 2, \ldots, n$, then inequality (4) simplifies to become

$$0 \leq 0 \times \sum_{j=1}^{n} s_j^2 = 0,$$

which is clearly true. Therefore, we need only consider cases where $(r_1, r_2, \ldots, r_n) \neq (0, 0, \ldots, 0)$.

Let $x \in \mathbb{R}$ and consider the expression

$$\sum_{j=1}^{n} (r_j x + s_j)^2.$$

Since a sum of squares is always non-negative, it follows that for any real number x,

$$\sum_{j=1}^{n} (r_j x + s_j)^2 \geq 0,$$

with equality if, and only if, each term in the sum is 0 (which means that $s_j = -xr_j$ for each j).

Multiplying out the left-hand side of this inequality, we obtain

$$\sum_{j=1}^{n}(r_j^2x^2 + 2r_js_jx + s_j^2) \geq 0,$$

or equivalently

$$\left(\sum_{j=1}^{n} r_j^2\right)x^2 + 2\left(\sum_{j=1}^{n} r_js_j\right)x + \left(\sum_{j=1}^{n} s_j^2\right) \geq 0.$$

The left-hand side is a quadratic expression $Ax^2 + Bx + C$ with

$$A = \sum_{j=1}^{n} r_j^2, \quad B = 2\sum_{j=1}^{n} r_js_j \quad \text{and} \quad C = \sum_{j=1}^{n} s_j^2.$$

A is not zero since $(r_1, r_2, \ldots, r_n) \neq (0, 0, \ldots, 0)$.

So we have, for each choice of $x \in \mathbb{R}$,

$$Ax^2 + Bx + C \geq 0,$$

with equality if, and only if, $s_j = -xr_j$ for each j. Hence

$$B^2 \leq 4AC.$$

Substituting for A, B and C in this inequality, we obtain

$$4\left(\sum_{j=1}^{n} r_js_j\right)^2 \leq 4\left(\sum_{j=1}^{n} r_j^2\right)\left(\sum_{j=1}^{n} s_j^2\right).$$

Dividing each side by 4 gives the required inequality. ■

If we take the square root of both sides of equation (4), we obtain the inequality

$$\sum_{j=1}^{n} r_js_j \leq \left(\sum_{j=1}^{n} r_j^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{n} s_j^2\right)^{\frac{1}{2}}. \tag{5}$$

This inequality is also known as the Cauchy–Schwarz Inequality.

We can now prove the Triangle Inequality (M3) for the Euclidean distance on $\mathbb{R}^n$.

Triangle Inequality

For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$, $d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$.

In the proof, we write the Triangle Inequality in a form that enables us to see that it follows from the Cauchy–Schwarz Inequality. To do this we use the formula for the Euclidean distance function. There is nothing very deep here, but the algebraic manipulation requires some care, as it involves sums of n real numbers and their squares.

Proof Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be in $\mathbb{R}^n$. For convenience, we introduce the abbreviations $\mathbf{r} = \mathbf{a} - \mathbf{b}$ and $\mathbf{s} = \mathbf{b} - \mathbf{c}$, so that $\mathbf{r} + \mathbf{s} = \mathbf{a} - \mathbf{c}$. Thus

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{a} - \mathbf{b}, \mathbf{0}) = d^{(n)}(\mathbf{r}, \mathbf{0}) = \left(\sum_{j=1}^{n} r_j^2 \right)^{\frac{1}{2}},$$

$$d^{(n)}(\mathbf{b}, \mathbf{c}) = d^{(n)}(\mathbf{b} - \mathbf{c}, \mathbf{0}) = d^{(n)}(\mathbf{s}, \mathbf{0}) = \left(\sum_{j=1}^{n} s_j^2 \right)^{\frac{1}{2}},$$

$$d^{(n)}(\mathbf{a}, \mathbf{c}) = d^{(n)}(\mathbf{a} - \mathbf{c}, \mathbf{0}) = d^{(n)}(\mathbf{r} + \mathbf{s}, \mathbf{0}) = \left(\sum_{j=1}^{n} (r_j + s_j)^2 \right)^{\frac{1}{2}}.$$

Squaring and expanding, we find that

We are aiming to obtain an expression to which we can apply the Cauchy–Schwarz Inequality.
We square to avoid the complication of having to deal with the square root.

$$\begin{aligned} \left(d^{(n)}(\mathbf{a}, \mathbf{c}) \right)^2 &= \sum_{j=1}^{n} (r_j + s_j)^2 \\ &= \left(\sum_{j=1}^{n} r_j^2 \right) + 2 \left(\sum_{j=1}^{n} r_j s_j \right) + \left(\sum_{j=1}^{n} s_j^2 \right). \end{aligned}$$

If we now use the Cauchy–Schwarz Inequality, equation (5), to find an upper bound for the middle term on the right-hand side, we obtain

$$\begin{aligned} \left(d^{(n)}(\mathbf{a}, \mathbf{c}) \right)^2 &\leq \left(\sum_{j=1}^{n} r_j^2 \right) + 2 \left(\sum_{j=1}^{n} r_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} s_j^2 \right)^{\frac{1}{2}} + \left(\sum_{j=1}^{n} s_j^2 \right) \\ &= \left(\left(\sum_{j=1}^{n} r_j^2 \right)^{\frac{1}{2}} + \left(\sum_{j=1}^{n} s_j^2 \right)^{\frac{1}{2}} \right)^2 \\ &= \left(d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c}) \right)^2. \end{aligned}$$

Taking square roots, we obtain

$$d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c}),$$

which completes the proof of the Triangle Inequality for the usual Euclidean distance in $\mathbb{R}^n$. ■

2 Introducing metric spaces

We now introduce the idea of a metric space, and show how this concept allows us to generalise the notion of continuity. We will then concentrate on looking at some examples of metric spaces and defer further discussion of continuous functions between metric spaces to the next chapter.

Subsection 2.1 shows how properties of the Euclidean distance function on $\mathbb{R}^n$ can be generalised to define a metric on any set to create a metric space.

You should not expect to have a firm grasp of the idea of a metric space by the end of this section: this will come as you see more examples of metric spaces in Chapter 15 and later chapters.

2.1 The definition of a metric space

Looking back at the definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ that we gave in Section 1, we see that it depends on:

- the Euclidean distance in $\mathbb{R}^n$, $d^{(n)}$
- the Euclidean distance in $\mathbb{R}^m$, $d^{(m)}$
- the definition of convergence for sequences in both $\mathbb{R}^n$ and $\mathbb{R}^m$.

Recall that if $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$, then $d^{(k)}(\mathbf{a}, \mathbf{b})$ is given by $\sqrt{\sum_{i=1}^{k}(b_i - a_i)^2}$.

Moreover, the definition of convergence for sequences also depends on the notion of distance. Thus a notion of distance (in both the domain and the codomain) is a fundamental part of our definition of continuity. Hence to define a notion of continuity for functions between arbitrary sets X and Y, we must first have effective notions of distance appropriate to the sets X and Y. An obvious strategy to do this is to try to find some properties of the Euclidean distance that we would expect any sensible notion of distance to possess.

In Proposition 1.3, we showed that the Euclidean distance function has three properties.

The *Euclidean distance function* $d^{(n)}\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ has the following properties.

For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$:

(M1) $d^{(n)}(\mathbf{a}, \mathbf{b}) \geq 0$, with equality holding if, and only if, $\mathbf{a} = \mathbf{b}$
(M2) $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{a})$
(M3) $d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$ (the *Triangle Inequality*).

Properties (M1)–(M3) do not use any special properties of the set $\mathbb{R}^n$, so it seems reasonable to say that for any set X, a function $d\colon X \times X \to \mathbb{R}$ is a *distance function* if it satisfies them. It turns out that these are indeed the appropriate properties of distance on which to base our definition. We call such a distance function a *metric*, and we call a set with such a function defined on it a *metric space*.

The word *metric* comes from the Greek word *μετρόν* (*metron*), meaning 'measure', an instrument for measuring.

Definition 2.1 *Metric*

Let X be a set. A **metric** on X is a function $d\colon X \times X \to \mathbb{R}$ that satisfies the following three conditions.

For each $a, b, c \in X$:

(M1) $d(a,b) \geq 0$, with equality holding if, and only if, $a = b$
(M2) $d(a,b) = d(b,a)$
(M3) $d(a,c) \leq d(a,b) + d(b,c)$ (the **Triangle Inequality**).

The set X, together with a metric d on X, is called a **metric space**, and is denoted by (X, d).

This definition was proposed by Maurice Fréchet, whose work is discussed in the *History Reader*.

Remarks

A 'point' may be nothing like a dot in the plane: X could be a set of functions and then a point in X would be a function.

1. Mathematicians usually refer to the members of the set X as **points**, to emphasise the analogy with the points of a line, a plane or three-dimensional space. Similarly, mathematicians usually refer to $d(a,b)$ as the **distance** between a and b. The theory of metric spaces is the study of those properties of sets of points that depend only on distance.
2. Condition (M1) says that distance is a non-negative quantity, and that the only point of a metric space that is at zero distance from a given point is that point itself.
3. Condition (M2) says that the distance from a point a to a point b is precisely the same as the distance from b to a. (The metric is *symmetric*. It doesn't matter whether you are going from a to b or from b to a – the distance is the same.)
4. Condition (M3) tells us that $d(a,c)$ gives the 'shortest' distance between a and c. For if we go directly from a to c, that gives a distance of $d(a,c)$. If we make a detour to b along the way, we must go a distance $d(a,b)$ to get to b, and then an additional distance $d(b,c)$ to get from b to c (Figure 2.1). Condition (M3) tells us that the total distance travelled must then be at least as great as the 'direct' distance $d(a,c)$.
5. Since our general definition is modelled on the corresponding properties of the Euclidean distance function $d^{(n)}$, it follows that $(\mathbb{R}^n, d^{(n)})$ is a metric space, for each $n \in \mathbb{N}$. It is known as **Euclidean n-space**. Furthermore, in the context of metric spaces, the Euclidean distance function $d^{(n)}$ is often referred to as the **Euclidean metric** for $\mathbb{R}^n$. These are our first examples of metric spaces.

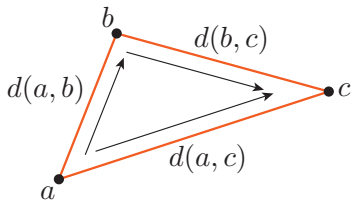


Figure 2.1 Shortest distance between a and c

If we look back at the proof of the Reverse Triangle Inequality for the Euclidean metric on $\mathbb{R}^n$ (Proposition 1.4), we see that the proof makes use only of properties (M1)–(M3) of the Euclidean metric. So it is no surprise to learn that there is a Reverse Triangle Inequality for any metric space.

Proposition 2.2 *Reverse Triangle Inequality for metric spaces*

Let (X, d) be a metric space. For each $a, b, c \in X$,
(M3a) $d(b, c) \geq |d(a, c) - d(a, b)|$.

The proof is essentially the same as that for Proposition 1.4, and so we omit it.

2.2 Examples of metric spaces

We now give a few examples of metric spaces with the aim of getting you used to showing that a distance function is a metric. The examples we consider here are not very sophisticated and may not persuade you of the utility of this new concept. However, in the next chapter we investigate ways of defining distances between more interesting 'points', such as between pairs of functions or pairs of sequences – this is when we will start to see the real power of metrics.

The discrete metric

Our first example is simple but important, since it shows that it is possible to define a metric on any set X.

Definition 2.3 *Discrete metric*

Let X be a set. The **discrete metric** on X is the function $d_0 \colon X \times X \to \mathbb{R}$ defined by

$$d_0(a, b) = \begin{cases} 0, & \text{if } a = b, \\ 1, & \text{if } a \neq b. \end{cases}$$

We show that d_0 is a metric after the remarks.

Remarks

1. When $X = \emptyset$, the empty set, the definition still makes sense but is not very interesting.
2. The d_0-distance between any two *distinct* points of a set X is always equal to 1. In particular, if we consider the case when $X = \mathbb{R}$, the real line, we find that d_0 is a different metric from the usual Euclidean distance. (To see this, observe that $d_0(0, 2) = 1$ whereas $d^{(1)}(0, 2) = 2$.) Thus the discrete metric gives us a second notion of distance on $\mathbb{R}$. In fact, if (X, d) is any metric space whose metric d is not d_0, then (X, d_0) is a second metric space with underlying set X.

   The discrete metric is about the simplest possible notion of distance there can be and acts to 'isolate' points of a space: they are all at distance 1 from one another.

Proposition 2.4

Let X be a set. Then (X, d_0) is a metric space.

Proof Let X be a set. In order to show that d_0 is a metric on X, we must show that d_0 satisfies conditions (M1)–(M3).

If $X = \emptyset$, then X has no elements and so each of (M1)–(M3) vacuously holds: there is nothing to prove. So suppose instead that X is non-empty.

(M1) Since d_0 can take only the values 0 and 1, we have $d_0(a, b) \geq 0$ for each choice of $a, b \in X$.

The definition of d_0 implies immediately that $d_0(a, b) = 0$ if, and only if, $a = b$.

Thus d_0 satisfies (M1).

(M2) If $a = b$, then $d_0(a, b) = 0 = d_0(b, a)$.

If $a \neq b$, then $d_0(a, b) = 1 = d_0(b, a)$.

Thus, for each $a, b \in X$, (M2) holds.

(M3) Let $a, b, c \in X$. We examine the two possible cases: $d_0(a, c) = 0$ and $d_0(a, c) = 1$.

Suppose $d_0(a, c) = 0$ (so $a = c$). Since $d_0(a, b)$ and $d_0(b, c)$ are non-negative, it follows that

$$d_0(a, b) + d_0(b, c) \geq 0 = d_0(a, c).$$

Now suppose $d_0(a, c) = 1$; then $a \neq c$ and so b cannot equal both a and c. Hence, from (M1), at least one of $d_0(a, b)$ and $d_0(b, c)$ is non-zero and so must equal 1. Thus

$$d_0(a, b) + d_0(b, c) \geq 1 = d_0(a, c).$$

Hence, in both cases, (M3) holds.

Since d_0 satisfies (M1)–(M3), it is a metric on X and (X, d_0) is a metric space. ■

At first appearance, the simplicity of the definition of the discrete metric d_0 may make it seem unimportant, but this is not the case. Its utility lies in the fact that it gives an 'extreme' example of a metric space – no other definition of distance so completely ignores any structure that may be present in the underlying set X.

We use the discrete metric for a number of purposes, principally to test properties that we suspect may hold for all metric spaces. If such a property fails for (X, d_0), then we know that our suspicion was false.

Exercise 2.1

Let $X = \{x, y, z\}$ and define $d\colon X \times X \to \mathbb{R}$ by

$$d(x,x) = d(y,y) = d(z,z) = 0,$$
$$d(x,y) = d(y,x) = 1,$$
$$d(y,z) = d(z,y) = 2,$$
$$d(x,z) = d(z,x) = 4.$$

Determine whether d is a metric on X.

The taxicab metric: an alternative metric for the plane

We have already observed that the usual definition of distance on the plane defines a metric and we have just seen that it is possible to define a very simple notion of distance on the plane, namely the discrete metric given previously. Here we introduce another natural way to measure distance between points in the plane, the taxicab metric.

Recall that the usual definition of distance on the plane is given by

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \quad \text{for } \mathbf{a}, \mathbf{b} \in \mathbb{R}^2.$$

What happens if we change the formula we use for distance on the right-hand side of this expression to something that is perhaps easier to calculate? For example, consider the function $e_1\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$e_1(\mathbf{a}, \mathbf{b}) = |b_1 - a_1| + |b_2 - a_2| \quad \text{for } \mathbf{a}, \mathbf{b} \in \mathbb{R}^2.$$

As we will shortly see, this *does* define an alternative metric on the plane. First let us understand how this distance function behaves.

The reason for the subscript 1 in the notation e_1 will become clear in Chapter 15.

Exercise 2.2

Find:

(a) $e_1((0,0),(1,0))$; (b) $e_1((0,0),(0,1))$;

(c) $e_1((0,1),(1,0))$.

You may have realised from your work on Exercise 2.2 that one way to understand $e_1(\mathbf{a}, \mathbf{b})$ is to draw a right-angled triangle with $\mathbf{a}$ and $\mathbf{b}$ at its two non-right-angled vertices, as shown in Figure 2.2. Then $e_1(\mathbf{a}, \mathbf{b})$ is the sum of the distances along the sides parallel to the axes. Another way to interpret it is to consider $\mathbf{a}$ and $\mathbf{b}$ as representing intersections in a city where the roads form a rectangular grid, as Figure 2.3 illustrates. Then $e_1(\mathbf{a}, \mathbf{b})$ is the shortest distance that can be travelled by a vehicle, such as a taxi, to get from $\mathbf{a}$ to $\mathbf{b}$. This interpretation leads to the common name *taxicab metric* for e_1.

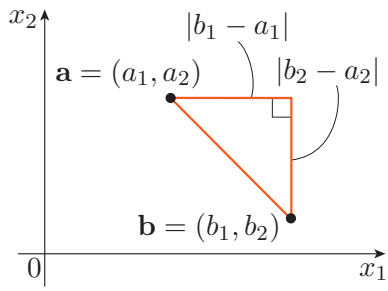


Figure 2.2 Visual representation of $e_1(\mathbf{a}, \mathbf{b})$

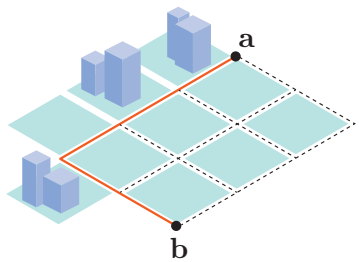


Figure 2.3 Alternative interpretation of $e_1(\mathbf{a}, \mathbf{b})$

Let us verify that e_1 does define a metric: we must check that it satisfies conditions (M1)–(M3).

(M1) For each $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$, $|b_1 - a_1|$ and $|b_2 - a_2|$ are non-negative, and hence so is their sum: thus $e_1(\mathbf{a}, \mathbf{b}) \geq 0$.

For each $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$,

$$e_1(\mathbf{a}, \mathbf{a}) = |a_1 - a_1| + |a_2 - a_2| = 0 + 0 = 0.$$

Conversely, suppose $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$ are such that $e_1(\mathbf{a}, \mathbf{b}) = 0$. Then

$$0 = |b_1 - a_1| + |b_2 - a_2|,$$

which implies that both $|b_1 - a_1| = 0$ and $|b_2 - a_2| = 0$. Thus $a_1 = b_1$ and $a_2 = b_2$; that is, $\mathbf{a} = \mathbf{b}$.

This proves that condition (M1) is satisfied.

The modulus function is the Euclidean distance function on $\mathbb{R}$, and so is a metric.

(M2) Let $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$. Using that the modulus function has property (M2), we obtain

$$\begin{aligned} e_1(\mathbf{a}, \mathbf{b}) &= |b_1 - a_1| + |b_2 - a_2| \\ &= |a_1 - b_1| + |a_2 - b_2| = e_1(\mathbf{b}, \mathbf{a}). \end{aligned}$$

This proves that condition (M2) is satisfied by e_1.

(M3) Let $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2), \mathbf{c} = (c_1, c_2) \in \mathbb{R}^2$. Then, using that the modulus function has property (M3),

$$\begin{aligned} e_1(\mathbf{a}, \mathbf{c}) &= |c_1 - a_1| + |c_2 - a_2| \\ &\leq (|c_1 - b_1| + |b_1 - a_1|) + (|c_2 - b_2| + |b_2 - a_2|) \\ &= (|b_1 - a_1| + |b_2 - a_2|) + (|c_1 - b_1| + |c_2 - b_2|) \\ &= e_1(\mathbf{a}, \mathbf{b}) + e_1(\mathbf{b}, \mathbf{c}). \end{aligned}$$

This proves that condition (M3) is satisfied by e_1.

We conclude that e_1 is a metric on the plane.

Exercise 2.3

Determine which of the following functions define a metric on $\mathbb{R}$:

(a) $d(a, b) = \sqrt{b^2 - a^2}$; (b) $d(a, b) = \sqrt{|b|^3 + |a|^3}$;

(c) $d(a, b) = \sqrt[3]{|b^3 - a^3|}$.

A 'mixed' metric in the plane

Our next example of a metric is a mixture of the discrete metric and the modulus function.

For points $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$ in the plane, define the function $d\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ by

$$d(\mathbf{a}, \mathbf{b}) = |b_1 - a_1| + d_0(a_2, b_2),$$

where d_0 denotes the discrete metric on $\mathbb{R}$.

Our goal is to show that this is indeed a metric on the plane. To do this, we make extensive use of the fact that the modulus function and the discrete metric both satisfy properties (M1)–(M3) on $\mathbb{R}$.

Exercise 2.4

Show that d has property (M1).

Exercise 2.5

Show that d has property (M2).

We verify the Triangle Inequality as a worked exercise.

Worked Exercise 2.5

Show that d has property (M3).

Solution

We must show that for any choice of $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^2$, we have

$$d(\mathbf{a}, \mathbf{c}) \leq d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}).$$

As with the previous exercises, we use the fact that both $|\cdot|$ and d_0 have property (M3) when viewed as metrics on $\mathbb{R}$.

Let $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, $\mathbf{c} = (c_1, c_2) \in \mathbb{R}^2$. Then

$$\begin{aligned} d(\mathbf{a}, \mathbf{c}) &= |a_1 - c_1| + d_0(c_2, a_2), \text{ by the definition of } d \\ &\leq |a_1 - b_1| + |b_1 - c_1| + d_0(c_2, a_2), \text{ by (M3) for } |\cdot| \\ &\leq |a_1 - b_1| + |b_1 - c_1| + d_0(c_2, b_2) + d_0(b_2, a_2), \text{ by (M3) for } d_0 \\ &= (|a_1 - b_1| + d_0(b_2, a_2)) + (|b_1 - c_1| + d_0(c_2, b_2)) \\ &= d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}), \end{aligned}$$

as required.

Since d has properties (M1)–(M3), we conclude that it is a metric for $\mathbb{R}^2$ but rather a strange one!

2.3 Understanding the geometry of metric spaces

Before we look at what it means for a sequence to be convergent with respect to a given metric, we spend a little time discussing one way of gaining some understanding about the geometric meaning of a given metric.

In the last subsection, we met three different metrics: the discrete metric, the taxicab metric on the plane and a mixed metric on the plane (which was formed from the usual distance in $\mathbb{R}$ together with the discrete metric).

An easy way to gain some insight into the behaviour of a metric is to look at the *balls* around a given point. For the usual Euclidean distance in $\mathbb{R}^n$, a ball of radius r around a point $\mathbf{a} \in \mathbb{R}^n$ consists of all those points whose distance from $\mathbf{a}$ is at most r, and this definition naturally extends to general metric spaces. However, in the following definition we take care to distinguish between balls that include points at exactly distance r from the centre $\mathbf{a}$ and those that do not.

Definition 2.6 *Open and closed balls*

Let (X, d) be a metric space, and let $a \in X$ and $r \geq 0$.

The **open ball** of **radius** r with **centre** a is the set

$$B_d(a, r) = \{x \in X : d(a, x) < r\}.$$

The **closed ball** of **radius** r with **centre** a is the set

$$B_d[a, r] = \{x \in X : d(a, x) \leq r\}.$$

The **sphere** of **radius** r with **centre** a is the set

$$S_d(a, r) = \{x \in X : d(a, x) = r\}.$$

When $r = 1$, these sets are called respectively the **unit open ball with centre** a, the **unit closed ball with centre** a and the **unit sphere with centre** a.

There is no universal agreement on the terminology and notation for balls. It is a wise precaution to check the conventions whenever reading a book on metric spaces.

Remarks

1. When we wish to emphasise the particular metric d on X, we refer to the *d-open ball*, the *d-closed ball* and the *d-sphere*.
2. When d is the usual Euclidean metric on $\mathbb{R}^2$, we recover the usual notions of open and closed balls in the plane.
3. Since $d(a, a) = 0$, $a \in B_d(a, r)$ whenever $r > 0$ and $a \in B_d[a, r]$ whenever $r \geq 0$.

Worked Exercise 2.7

Let (X, d) be a metric space, and let $a \in X$. Show that

$$B_d(a, 0) = \emptyset, \quad B_d[a, 0] = \{a\} \quad \text{and} \quad S_d(a, 0) = \{a\}.$$

Solution

It follows from (M1) that

$$B_d(a, 0) = \{x \in X : d(a, x) < 0\} = \emptyset,$$

$$\begin{aligned} B_d[a, 0] &= \{x \in X : d(a, x) \leq 0\} \\ &= \{x \in X : d(a, x) = 0\} = \{a\}, \end{aligned}$$

and

$$S_d(a, 0) = \{x \in X : d(a, x) = 0\} = \{a\}.$$

We now discover what open balls, closed balls and spheres look like for some of the metric spaces we have met already.

Let us start by determining the open and closed balls for the discrete metric, d_0.

Worked Exercise 2.8

Let X be a non-empty set and $a \in X$. Determine $B_{d_0}(a, r)$ for $r \geq 0$.

Solution

Let $a \in X$ and suppose that $r > 0$.

Since $B_{d_0}(a, r) = \{x \in X : d_0(a, x) < r\}$ and $d_0(a, x) = 1$ unless $a = x$ (when it is 0), we conclude that

$$B_{d_0}(a, r) = \begin{cases} \emptyset, & \text{if } r = 0, \\ \{a\}, & \text{if } 0 < r \leq 1, \\ X, & \text{if } r > 1. \end{cases}$$

Exercise 2.6

Let X be a non-empty set and $a \in X$. Determine $B_{d_0}[a, r]$ for $r \geq 0$.

Exercise 2.7

Let X be a non-empty set and $a \in X$. Determine $S_{d_0}(a, r)$ for $r \geq 0$.

Next, let us look at open balls defined using the taxicab metric.

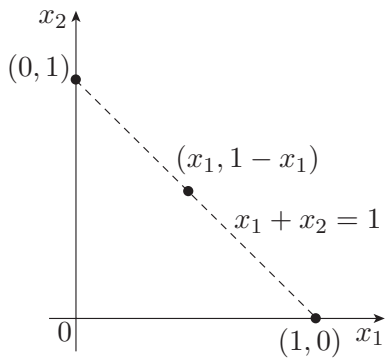


Figure 2.4 The line $x_1 + x_2 = 1$

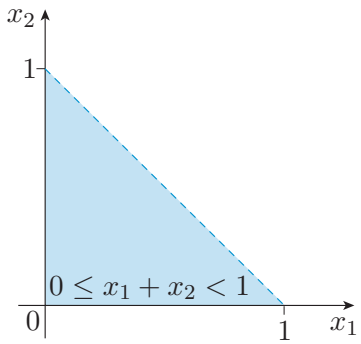


Figure 2.5 The points where $x_1 + x_2 < 1$

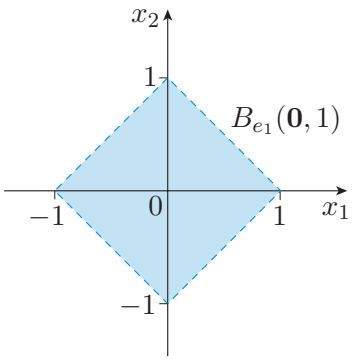


Figure 2.6 The open ball $B_{e_1}(\mathbf{0}, 1)$

Worked Exercise 2.9

Consider the metric space $(\mathbb{R}^2, e_1)$ – that is, the plane with the taxicab metric. Find the unit open ball $B_{e_1}(\mathbf{0}, 1)$.

Solution

The centre is $\mathbf{0} = (0, 0)$, and we want to find all points $\mathbf{x} = (x_1, x_2)$ that satisfy

$$e_1(\mathbf{0}, \mathbf{x}) = |x_1| + |x_2| < 1.$$

We first consider points in the first quadrant, where $x_1, x_2 \geq 0$.

We want to find those points where $x_1 + x_2 < 1$. Consider the line $x_1 + x_2 = 1$, or equivalently $x_2 = 1 - x_1$. In the first quadrant, this line connects the points $(0, 1)$ and $(1, 0)$ and is shown dashed in Figure 2.4. The points on this line segment have coordinates $(x_1, 1 - x_1)$. All points below the line segment have coordinates (x_1, x_2) with $x_2 < 1 - x_1$ and all points on or above it have coordinates (x_1, x_2) with $x_2 \geq 1 - x_1$. Hence the points where $x_1 + x_2 < 1$ are those strictly below the line segment, making up the shaded region in Figure 2.5.

By use of a similar argument for each of the other three quadrants, or by appealing to the symmetry of the situation, we obtain triangular regions in each quadrant. Combining these, we obtain the diamond-shaped region in Figure 2.6; the open ball $B_{e_1}(\mathbf{0}, 1)$ is the set of points strictly inside this diamond, shown shaded in the figure. The dashed boundary indicates that it is not included in the set.

Exercise 2.8

Sketch the open ball $B_{e_1}((2, 3), 2)$, briefly justifying your answer.

Let us finish this discussion by looking at the balls for our mixed metric in the plane, d.

Worked Exercise 2.10

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$. Determine $B_d(\mathbf{a}, r)$ for $r > 0$, where d denotes the mixed metric in the plane given by

$$d(\mathbf{a}, \mathbf{b}) = |b_1 - a_1| + d_0(a_2, b_2) \text{ for } \mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2.$$

Solution

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$ and suppose that $r > 0$. Then, using the definition of an open ball,

$$\begin{aligned} B_d(\mathbf{a}, r) &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : d(\mathbf{a}, \mathbf{x}) < r\} \\ &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : |x_1 - a_1| + d_0(a_2, x_2) < r\}. \end{aligned}$$

We now consider the cases $x_2 = a_2$ and $x_2 \neq a_2$ separately. If $x_2 = a_2$, then $d_0(a_2, x_2) = 0$ and so

$$\begin{aligned} \{\mathbf{x} = (x_1, a_2) \in \mathbb{R}^2 : |x_1 - a_1| < r\} &= \{x_1 \in \mathbb{R} : |x_1 - a_1| < r\} \times \{a_2\} \\ &= (a_1 - r, a_1 + r) \times \{a_2\}. \end{aligned}$$

If $x_2 \neq a_2$, then $d_0(a_2, x_2) = 1$ and

$$\begin{aligned} &\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : x_2 \neq a_2 \text{ and } |x_1 - a_1| + d_0(a_2, x_2) < r\} \\ &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : x_2 \neq a_2 \text{ and } |x_1 - a_1| < r - 1\}. \end{aligned}$$

Let us unpack this set expression: x_2 can be any value other than a_2 but if $r \leq 1$, then there are no possible x_1 for which $|x_1 - a_1| < r - 1$, and so the expression is the empty set, $\emptyset$. Whereas if $r > 1$, then the set of $x_1 \in \mathbb{R}$ for which $|x_1 - a_1| < r - 1$ is the interval $(a_1 - (r-1), a_1 + (r-1))$. Hence for $r > 1$, the set expression is $(a_1 - (r-1), a_1 + (r-1)) \times (\mathbb{R} - \{a_2\})$. Summarising,

$$\begin{aligned} &\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : x_2 \neq a_2 \text{ and } |x_1 - a_1| + d_0(a_2, x_2) < r\} \\ &= \begin{cases} \emptyset, & \text{if } r \leq 1, \\ (a_1 - (r-1), a_1 + (r-1)) \times (\mathbb{R} - \{a_2\}), & \text{if } r > 1. \end{cases} \end{aligned}$$

Hence, as shown in Figure 2.7,

$$B_d(\mathbf{a}, r) = \begin{cases} (a_1 - r, a_1 + r) \times \{a_2\}, & \text{if } r \leq 1, \\ ((a_1 - r, a_1 + r) \times \{a_2\}) & \\ \qquad \cup ((a_1 - (r-1), a_1 + (r-1)) \times \mathbb{R}), & \text{if } r > 1. \end{cases}$$

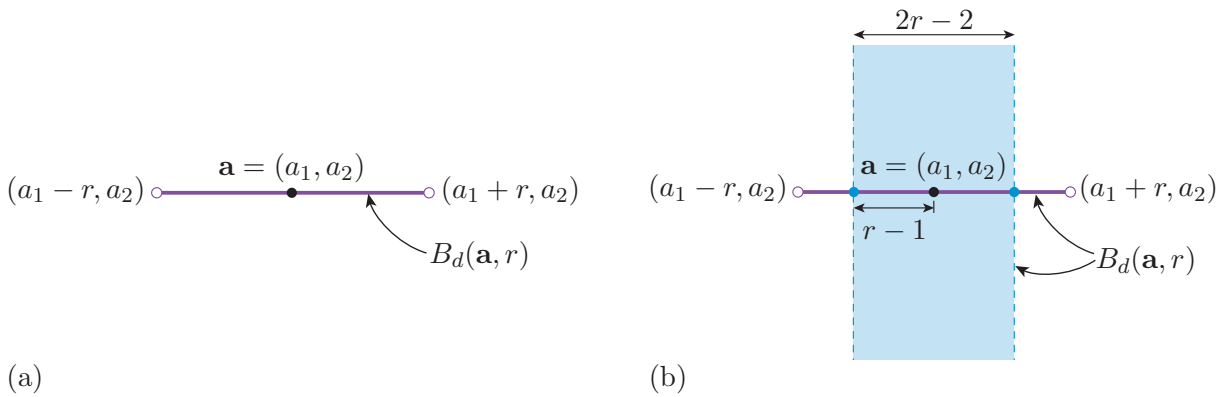


Figure 2.7 $B_d(\mathbf{a}, r)$: (a) for $r \leq 1$; (b) for $r > 1$

Exercise 2.9

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$. Determine $B_d[\mathbf{a}, r]$ for $r \geq 0$, where d denotes the mixed metric in the plane given by

$$d(\mathbf{a}, \mathbf{b}) = |b_1 - a_1| + d_0(a_2, b_2) \text{ for } \mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2.$$

We will return to our investigation of the geometry of metric spaces in Chapter 16.

3 Sequences in metric spaces

Now that we have several examples of metric spaces available to us, we return to the problem of defining continuous functions between metric spaces.

Since the definition of a general metric space is modelled on the properties of the Euclidean metric $d^{(n)}$ on $\mathbb{R}^n$, and we defined continuity of functions between Euclidean spaces in terms of convergent sequences, it is natural to attempt to extend our ideas about convergent sequences in $\mathbb{R}^n$ to general metric spaces. In fact, we did much of the hard work when we generalised from the notion of convergence for real-valued sequences to that of convergence of sequences in $\mathbb{R}^n$; it is now only a short step to develop these concepts for the metric space setting.

In Chapter 13, we observed that a real sequence can be thought of as a function $a\colon \mathbb{N} \to \mathbb{R}$, given by $n \mapsto a_n$. Note that the only role played by $\mathbb{R}$ here is as the codomain of the function $a\colon \mathbb{N} \to \mathbb{R}$; the structure of $\mathbb{R}$ becomes relevant only when convergence is considered. Since the codomain of a function is simply a set, the following definition is a natural generalisation.

Definition 3.1 *Sequence in a metric space*

Let X be a set. A **sequence** in X is an unending ordered list of elements of X:

$$a_1, a_2, a_3, \ldots.$$

The element a_k is the ***k*th term** of the sequence, and the whole sequence is denoted by (a_k), $(a_k)_{k=1}^{\infty}$ or $(a_k)_{k\in\mathbb{N}}$.

Note that this definition of a sequence does not require that we impose any additional structure (such as a metric) on the set X.

The definition of what it means for a sequence to converge in a metric space (X, d) is closely based on the definition of convergence in $\mathbb{R}^n$ given in Definition 1.6.

Definition 3.2 *Convergence in a metric space*

Let (X, d) be a metric space. A sequence (a_k) in X **d-converges** to $a \in X$ if the sequence of real numbers $(d(a_k, a))$ is a null sequence.

We write $a_k \xrightarrow{d} a$ as $k \to \infty$, or simply $a_k \to a$ if the context is clear.

We say that the sequence (a_k) is **convergent** in (X, d) with **limit** a.

A sequence that does not converge (with respect to the metric d) to any point in X is said to be **d-divergent**.

Remarks

1. Note that the limit a *must* be a point in X. For example, if X is the *open* interval $(0, 1)$ and we give it the usual notion of distance $d(a, b) = |a - b|$, then the sequence $(\frac{1}{n})$ is not d-convergent, since the only possible choice of limit is 0 – which is *not* a point of X.
2. In order to show that a sequence (a_k) converges to a for the metric d, we must show that $d(a_k, a) \to 0$ as $k \to \infty$ – that is, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ for which $d(a_k, a) < \varepsilon$ for all $k > N$.

It is not hard to check that this defines a metric on $(0, 1)$ – we return to the general problem of defining metrics on subsets of metric spaces in the next chapter.

Exercise 3.1

Let $(\mathbb{R}^2, e_1)$ be the plane with the taxicab metric, and let $(\mathbf{a}_n)$ be the sequence given by $\mathbf{a}_k = (1 + \frac{1}{2k}, 2 - \frac{1}{2k})$. Show that $(\mathbf{a}_k)$ converges to $(1, 2)$ with respect to e_1.

We saw in Lemma 1.7 that convergent sequences in $(\mathbb{R}^n, d^{(n)})$ have unique limits – that is, a sequence cannot simultaneously converge to two different limits. The next result establishes this as a fact in any metric space.

Theorem 3.3 *Uniqueness of limits in a metric space*

Let (X, d) be a metric space and let $a, b \in X$. If (a_k) is a sequence in X that d-converges to both a and b, then $a = b$.

Proof We use proof by contradiction.

Suppose that the sequence (a_k) d-converges to both a and b in X, with $a \neq b$. Then by property (M1) of Definition 2.1 for d, $d(a,b) > 0$ and so if we let $\varepsilon = \frac{1}{2}d(a,b)$, then $\varepsilon > 0$.

Since we are supposing that the sequence (a_k) converges to both a and b, the sequences of real numbers $(d(a_k,a))$ and $(d(a_k,b))$ are both null. Hence there is $N \in \mathbb{N}$ for which $d(a_k,a) < \varepsilon$ and $d(a_k,b) < \varepsilon$ whenever $k > N$.

The Triangle Inequality (property (M3) for d) tells us that, for each $k > N$,

$$\begin{aligned} d(a,b) \leq d(a,a_k) + d(a_k,b) &< \varepsilon + \varepsilon = 2\varepsilon \\ &= d(a,b), \end{aligned}$$

by the definition of ε. But this is impossible; hence our initial assumption that a sequence could converge to two distinct limits must be wrong. We conclude that any d-convergent sequence has a unique limit. ■

The next definition enables us to describe a particularly straightforward type of convergent sequence.

Definition 3.4 *Eventually constant*

Let (a_k) be a sequence in a set X. We say that (a_k) is **eventually constant** if there is $a \in X$ and $N \in \mathbb{N}$ such that $a_k = a$ whenever $k > N$.

Worked Exercise 3.5

Let (X,d) be a metric space and let (a_k) be a sequence in X that is eventually constant. Show that (a_k) converges for d and state its limit.

Solution

Since (a_k) is eventually constant, there is $a \in X$ and $N \in \mathbb{N}$ such that $a_k = a$ whenever $k > N$. Therefore

$$d(a_k,a) = d(a,a) = 0, \quad \text{for each } k > N.$$

Hence the sequence of real numbers $(d(a_n,a))$ is null, which means that $a_k \xrightarrow{d} a$ as $k \to \infty$. Thus the sequence (a_k) is d-convergent with limit a.

Convergence and the discrete metric

Let us consider the sequence (a_k) in $(\mathbb{R}, d_0)$ given by $a_k = \frac{1}{k}$. We see that

$$d_0(a_k, 0) = d_0\left(\tfrac{1}{k}, 0\right) = 1, \quad \text{for each } k \in \mathbb{N},$$

so (a_k) does not converge to 0 with respect to the discrete metric d_0. In fact, this sequence (a_k) is divergent for d_0, because given any non-zero $l \in \mathbb{R}$,

$$d_0(a_k, l) = d_0(\tfrac{1}{k}, l) = 1 \quad \text{for each } k > \tfrac{1}{|l|}.$$

$(\mathbb{R}, d_0)$ is the set of real numbers with the discrete metric d_0.

On the other hand, we know that the sequence $(\frac{1}{k})$ *does* converge to 0 with respect to the Euclidean metric $d^{(1)}$ on $\mathbb{R}$. The important conclusion to draw from this example is:

convergence depends on how we measure distance.

In other words, different metrics on the same set can give rise to different convergent sequences.

From the solution to Worked Exercise 3.5, we know that eventually constant sequences (a_k) in a metric space (X, d_0) are d_0-convergent. The next exercise asks you to show that the eventually constant sequences are the *only* convergent sequences for the d_0 metric on a set X.

Exercise 3.2

Let X be a set and d_0 the discrete metric for X. Suppose that (a_k) is a sequence in X that is d_0-convergent. Show that (a_k) is an eventually constant sequence.

4 The definition of continuity in metric spaces

Now that we know what it means for a sequence to converge in a metric space, we can formulate a definition of continuity for functions between metric spaces.

In Subsection 1.4 we discussed the definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ (Definition 1.11). The only property of the domain and the codomain used in that definition is the fact that we can identify convergent sequences by using the metrics $d^{(n)}$ and $d^{(m)}$. In consequence, we now generalise to give a definition of continuity for functions between *any* pair of metric spaces.

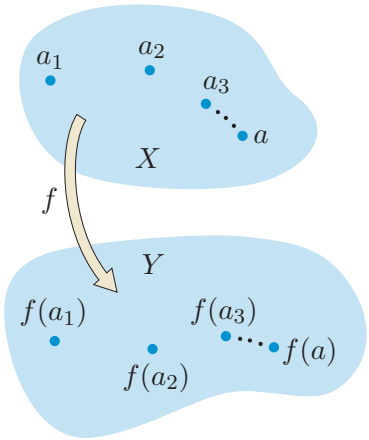


Figure 4.1 The function $f: X \to Y$

Definition 4.1 *Continuity for metric spaces*

Let (X, d) and (Y, e) be metric spaces and let $f: X \to Y$ be a function (Figure 4.1).

Then f is (d, e)-**continuous** at $a \in X$ if:

> whenever (a_k) is a sequence in X for which $a_k \xrightarrow{d} a$ as $k \to \infty$,
>
> then the sequence $f(a_k) \xrightarrow{e} f(a)$ as $k \to \infty$.

If f does not satisfy this condition at some $a \in X$ – that is, there is a sequence (x_k) in X for which $x_k \to a$ as $k \to \infty$ but $f(x_k)$ does not converge to $f(a)$ – then we say that f is (d, e)-**discontinuous** at a.

A function that is continuous at all points of X is said to be (d, e)-**continuous** on X (or simply **continuous**, if no ambiguity is possible).

Remarks

1. When $(X, d) = (\mathbb{R}^n, d^{(n)})$ and $(Y, e) = (\mathbb{R}^m, d^{(m)})$, we recover the definition of continuity given in Definition 1.11 of Subsection 1.4.
2. We have not yet defined what it means to be continuous on a subset A of the metric space X; we return to this issue in the next chapter.

See Theorem 1.6 of Chapter 15.

Our next worked exercise shows that this definition can make some surprising functions continuous.

Worked Exercise 4.2

Let $f: \mathbb{R} \to \mathbb{R}$ be a function and let $a \in \mathbb{R}$. Prove that f is always $(d_0, d^{(1)})$-continuous at a.

Solution

Let $a \in \mathbb{R}$ and suppose that (x_k) is a sequence in $\mathbb{R}$ that is d_0-convergent to a.

Then by Exercise 3.2 we deduce that there is $N \in \mathbb{N}$ so that for $k > N$, $x_k = a$. But then for $k > N$, $f(x_k) = f(a)$ and so for such k, $d^{(1)}(f(x_k), f(a)) = |f(a) - f(a)| = 0$. Hence $(d^{(1)}(f(x_k), f(a)))$ is a real null sequence and we conclude that f is $(d_0, d^{(1)})$-continuous at a.

This is a rather artificial example and it tells us that every function from $\mathbb{R}$ to $\mathbb{R}$ is $(d_0, d^{(1)})$-continuous on $\mathbb{R}$. However, it does illustrate that our intuitive notion of what continuity means breaks down when looking at metrics different from the Euclidean ones, and so highlights the importance of working from the definition.

In fact one can use a similar argument to show that any $f: X \to Y$ is (d_0, e)-continuous on X, no matter the choice of metric e for Y.

Worked Exercise 4.3

Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x_1, x_2) = (x_1, 2x_2)$. Prove that f is $(d^{(2)}, e_1)$-continuous on $\mathbb{R}^2$.

Recall that e_1 denotes the taxicab metric on the plane and $d^{(2)}$ is the usual Euclidean distance on the plane.

Solution

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$. We must show that if $(\mathbf{x}_k)$ is a sequence in the plane that $d^{(2)}$-converges to $\mathbf{a}$, then $(e_1(f(\mathbf{x}_k), f(\mathbf{a})))$ is a real null sequence.

Suppose that $(\mathbf{x}_k = (x_{1,k}x_{2,k}))$ is a sequence in the plane that $d^{(2)}$-converges to $\mathbf{a}$, that is, a sequence for which $(d^{(2)}(\mathbf{x}_k, \mathbf{a}))$ is a real null sequence. Then

$$\begin{aligned} e_1(f(\mathbf{x}_k), f(\mathbf{a})) &= |x_{1,k} - a_1| + |2x_{2,k} - 2a_2|, \text{ by definition of } f \text{ and } e_1 \\ &= |x_{1,k} - a_1| + 2|x_{2,k} - a_2| \\ &\leq 2(|x_{1,k} - a_1| + |x_{2,k} - a_2|) \\ &\leq 2(d^{(2)}(\mathbf{x}_k, \mathbf{a}) + d^{(2)}(\mathbf{x}_k, \mathbf{a})) = 4d^{(2)}(\mathbf{x}_k, \mathbf{a}). \end{aligned}$$

But $e_1(f(\mathbf{x}_k), f(\mathbf{a})) \geq 0$ for every k and we are assuming that $(d^{(2)}(\mathbf{x}_k, \mathbf{a}))$ is a real null sequence. Hence by the Squeeze Rule, $e_1(f(\mathbf{x}_k), f(\mathbf{a})) \to 0$ as $k \to \infty$. That is, $(e_1(f(\mathbf{x}_k), f(\mathbf{a})))$ is also a real null sequence and so f is $(d^{(2)}, e_1)$-continuous at a.

Since $a \in \mathbb{R}^2$ was an arbitrary point in $\mathbb{R}^2$, we conclude that f is $(d^{(2)}, e_1)$-continuous on $\mathbb{R}^2$.

At the moment our stock of metric spaces is quite small: Euclidean spaces, the plane with the taxicab metric, the plane with a particular 'mixed' metric, and arbitrary sets with the discrete metric. In the next chapter we will look at more examples of metric spaces and examine further the notion of continuity. What we can do at this point, though, is prove a useful result that applies to all continuous functions and which is an extension of the Composition Rule for continuous functions between Euclidean spaces.

Proposition 4.4 *Composition Rule*

Let (X, d_X), (Y, d_Y) and (Z, d_Z) be metric spaces. Let $f\colon X \to Y$ be (d_X, d_Y)-continuous and let $g\colon Y \to Z$ be (d_Y, d_Z)-continuous. Then the composed function $g \circ f\colon X \to Z$ is (d_X, d_Z)-continuous.

Recall that $(g \circ f)(x) = g(f(x))$.

Proof We take an arbitrary point $a \in X$ and prove that $g \circ f$ is continuous at a.

We need to show that if (a_k) is a sequence in X with $a_k \xrightarrow{d_X} a$ as $k \to \infty$, then

$$(g \circ f)(a_k) \xrightarrow{d_Z} (g \circ f)(a). \tag{1}$$

Write $y_k = f(a_k)$ and $y = f(a)$. Since f is (d_X, d_Y)-continuous on X, we know that

$$y_k = f(a_k) \xrightarrow{d_Y} f(a) = y \text{ as } k \to \infty.$$

That is, the sequence $(d_Y(y_k, y))$ is a real null sequence.

But the function g is (d_Y, d_Z)-continuous on Y and so it follows by the definition of continuity for g that

$$g(y_k) \xrightarrow{d_Z} g(y) \text{ as } k \to \infty.$$

This means that equation (1) holds, because $g(y_k) = (g \circ f)(a_k)$ and $g(y) = (g \circ f)(a)$.

Thus $g \circ f$ is continuous at the arbitrary point $a \in X$, and so we conclude that $g \circ f$ is continuous on X. ■

Exercise 4.1

Let (X, d) and (Y, e) be metric spaces, and let $b \in Y$ be fixed. Suppose that $f\colon X \to Y$ is (d, e)-continuous on X.

You may find parts of the method of Worked Exercise 1.15 helpful for Exercise 4.1.

Use the Reverse Triangle Inequality to prove that $g\colon X \to \mathbb{R}$ given by $g(x) = e(f(x), b)$ is $(d, d^{(1)})$-continuous on X.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should be able to:

(a) Understand the Euclidean distance function on $\mathbb{R}^n$, and appreciate its properties.

(b) State and use the Triangle and Reverse Triangle Inequalities for the Euclidean distance function on $\mathbb{R}^n$.

(c) Explain the definition of continuity for functions from $\mathbb{R}^n$ to $\mathbb{R}^m$.

(d) Determine whether a given function from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous.

(e) Explain the geometric meaning of each of the metric space properties (M1)–(M3) and be able to verify whether a given distance function is a metric.

(f) Distinguish between open and closed balls in a metric space and be able to determine them for given metric spaces.

(g) Define convergence for sequences in a metric space.

(h) Determine whether a given sequence in a metric space converges.

(i) State the definition of continuity of a function between two metric spaces.

Solutions to exercises

Solution to Exercise 1.1

In line with the remarks following Definition 1.1, we expect $A_1 \times A_2$ to have $2 \times 4 = 8$ elements and, indeed,

$$A_1 \times A_2 = \{(a,0),(a,1),(a,2),(a,3),(b,0),(b,1),(b,2),(b,3)\},$$

an eight-element set.

Solution to Exercise 1.2

$$\begin{aligned} &d^{(5)}((4,2,3,0,-1),(2,-3,-1,0,2)) \\ &\quad = \sqrt{(-2)^2+(-5)^2+(-4)^2+0^2+3^2} \\ &\quad = \sqrt{4+25+16+0+9} = \sqrt{54} = 3\sqrt{6}. \end{aligned}$$

Solution to Exercise 1.3

(M2) holds since

$$\begin{aligned} d^{(n)}(\mathbf{a},\mathbf{b}) &= \sqrt{(b_1-a_1)^2+(b_2-a_2)^2+\cdots+(b_n-a_n)^2}, \\ d^{(n)}(\mathbf{b},\mathbf{a}) &= \sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_n-b_n)^2}. \end{aligned}$$

In the two expressions, corresponding terms inside the square root are equal, since $(p-q)^2 = (q-p)^2$ for any $p, q \in \mathbb{R}$. Hence $d^{(n)}(\mathbf{a},\mathbf{b}) = d^{(n)}(\mathbf{b},\mathbf{a})$.

Solution to Exercise 1.4

(a) We recall that $\cos k\pi = (-1)^k$ and $\sin k\pi = 0$ for $k \in \mathbb{N}$, and so

$$\begin{aligned} \mathbf{a}_k &= \left(\frac{-1}{(-3)^k}, \frac{3^k + 5\times(-1)^k}{3^k}, 0, -1 - \frac{2}{3^k}\right) \\ &= \left(\frac{-1}{(-3)^k}, 1 + \frac{5\times(-1)^k}{3^k}, 0, -1 - \frac{2}{3^k}\right). \end{aligned}$$

Based on this, we guess that $(\mathbf{a}_k)$ may converge to $(0,1,0,-1)$. We calculate

$$\begin{aligned} &d^{(4)}(\mathbf{a}_k,(0,1,0,-1)) \\ &= \sqrt{\left(\frac{1}{(-3)^k}\right)^2 + \left(-\frac{5\times(-1)^k}{3^k}\right)^2 + 0^2 + \left(\frac{2}{3^k}\right)^2} \\ &= \sqrt{\frac{1}{3^{2k}} + \frac{25}{3^{2k}} + \frac{4}{3^{2k}}} = \frac{\sqrt{30}}{3^k}. \end{aligned}$$

We recognise $\left(\frac{\sqrt{30}}{3^k}\right)$ as a multiple of a basic null sequence, and hence deduce that $(\mathbf{a}_k)$ does indeed converge to $(0,1,0,-1)$.

(b) Let $(\mathbf{b}_k)$ be a constant sequence in $\mathbb{R}^n$. Then $\mathbf{b}_k = \mathbf{b}_1$ for each $k \in \mathbb{N}$. Hence the distance $d^{(n)}(\mathbf{b}_k, \mathbf{b}_1) = 0$ for each $k \in \mathbb{N}$ by property (M1) of the Euclidean distance $d^{(n)}$ (see Proposition 1.3).

Therefore, the real sequence $\big(d^{(n)}(\mathbf{b}_k, \mathbf{b}_1)\big) = (0)$ is null and thus $(\mathbf{b}_k)$ converges to $\mathbf{b}_1$.

Solution to Exercise 1.5

Since $\mathbf{a}_k \to \boldsymbol{\ell}$ as $k \to \infty$, Proposition 1.9 implies that for each $j = 1, 2, \ldots, n$, the real sequence $a_{j,k}$ converges to ℓ_j. But the Multiple Rule for real sequences implies that $\lambda a_{j,k} \to \lambda \ell_j$ as $k \to \infty$ for each $j = 1, 2, \ldots, n$. Hence, applying Proposition 1.9 again, we conclude $\lambda \mathbf{a}_k \to \lambda \boldsymbol{\ell}$ as $k \to \infty$.

Solution to Exercise 1.6

In view of Proposition 1.9, we just need to show that each of the three real sequences

$$(a_{1,k}) = \left(\frac{k^3 + 3^k}{k!}\right), \quad (a_{2,k}) = \left(1 - \frac{1}{\sqrt{k}}\right), \quad (a_{3,k}) = \big(2^{-k}(5 - 2^k)\big)$$

converges as $k \to \infty$.

Since $(a_{1,k})$ is the sum of the basic null sequences $(k^3/k!)$ and $(3^k/k!)$, the Sum Rule gives that $(a_{1,k})$ converges to 0.

By noting that $|a_{2,k} - 1| = \frac{1}{\sqrt{k}}$, and that $\left(\frac{1}{\sqrt{k}}\right)$ is a basic null sequence, we have that $(a_{2,k})$ converges to 1.

We observe that $a_{3,k} = 5(2^{-k}) - 1$, and so $|a_{3,k} - (-1)| = 5(2^{-k})$. Since $\left(\frac{5}{2^k}\right)$ is a null sequence, it follows that $(a_{3,k})$ converges to -1.

Proposition 1.9 now leads us to conclude that $(\mathbf{a}_k)$ converges to $(0, 1, -1)$.

Solution to Exercise 1.7

Let us first suppose that $(\mathbf{a}_k)$ converges to $\boldsymbol{\ell}$ and show that $a_{j,k} \to \ell_j$ for each $j = 1, 2, \ldots, n$. This means we need to show that the real sequences $(a_{j,k} - \ell_j)$ are null for each j. We estimate

$$\begin{aligned} |a_{j,k} - \ell_j| &= \sqrt{(a_{j,k} - \ell_j)^2} \\ &\leq \sqrt{(a_{1,k} - \ell_1)^2 + \cdots + (a_{n,k} - \ell_n)^2} = d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell}). \end{aligned}$$

We have assumed that $(\mathbf{a}_k)$ converges to $\boldsymbol{\ell}$, and so the real sequence $(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell}))$ must be null. Thus, applying the Squeeze Rule for real sequences to the above estimate shows that $(a_{j,k} - \ell_j)$ is null. Therefore $a_{j,k} \to \ell_j$ for each j.

See Chapter 13, Lemma 1.15

To complete the proof, we now suppose that $a_{j,k} \to \ell_j$ for $j = 1, \ldots, n$, and deduce from this that $\mathbf{a}_k \to \boldsymbol{\ell}$.

So, given that $(a_{j,k} - \ell_j)$ is a null sequence for each j, we use the Product and Sum Rules (for sequences of real numbers) to deduce that $\big((a_{1,k} - \ell_1)^2 + \cdots + (a_{n,k} - \ell_n)^2\big)$ is a null sequence of non-negative real numbers. Then the Power Rule for real null sequences, Theorem 1.14 in Chapter 13, gives that the sequence

$$(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})) = \left(\sqrt{(a_{1,k} - \ell_1)^2 + \cdots + (a_{n,k} - \ell_n)^2}\right)$$

is null. We conclude that $(\mathbf{a}_k)$ converges to $\boldsymbol{\ell}$ as $k \to \infty$, as required.

Solution to Exercise 1.8

By Proposition 1.9, the sequence $(\mathbf{b}_k)$ converges if, and only if, each of the coordinate sequences $(b_{j,k})$ converges for $j = 1, 2, 3, 4$. However, the sequence $(b_{2,k}) = (k)$ was shown to be divergent in Worked Exercise 1.18(b) of Chapter 13. Therefore, by Proposition 1.9, $(\mathbf{b}_k)$ does not converge.

(Note that we do not need to consider the convergence of the other coordinate sequences, but $\frac{k!}{k^k} \to 0$ and $k^{1/\log k} \to e$ as $k \to \infty$.)

Solution to Exercise 1.9

Let $\mathbf{a}$ be a general point in $\mathbb{R}^n$ and let $(\mathbf{x}_k)$ be a sequence in $\mathbb{R}^n$ such that $\mathbf{x}_k \to \mathbf{a}$. Then the sequence $f(\mathbf{x}_k)$ converges to $f(\mathbf{a})$ because the sequence of distances

$$\Big(d^{(m)}(f(\mathbf{x}_k), f(\mathbf{a}))\Big) = \Big(d^{(m)}(\mathbf{c}, \mathbf{c})\Big) = (0)$$

is clearly null. Hence f is continuous at $\mathbf{a}$. Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^n$, f is continuous on $\mathbb{R}^n$.

Solution to Exercise 1.10

The function g can be written as the composition of two functions that we already know to be continuous. Worked Exercise 1.15 with $\mathbf{u} = \mathbf{0}$ shows that $\mathbf{x} \mapsto d^{(n)}(\mathbf{x}, \mathbf{0})$ is continuous with image set $[0, \infty)$. The square function is one of the basic continuous functions from $\mathbb{R}$ to $\mathbb{R}$. We deduce, by the Composition Rule for continuous functions, that g is continuous on $\mathbb{R}^n$.

Solution to Exercise 1.11

Let $\mathbf{a}$ be a general point in $\mathbb{R}^n$ and let $(\mathbf{x}_k)$ be a sequence in $\mathbb{R}^n$ that converges to $\mathbf{a}$. Then, by Proposition 1.9,

$$p_j(\mathbf{x}_k) = x_{j,k} \to a_j = p_j(\mathbf{a}), \quad \text{as } k \to \infty.$$

Therefore p_j is continuous at $\mathbf{a}$.

Since $\mathbf{a}$ is an arbitrary point of $\mathbb{R}^n$, p_j is continuous on $\mathbb{R}^n$.

Solution to Exercise 1.12

We consider the composition of f with the three projection functions p_1, p_2 and p_3 from $\mathbb{R}^3$ to $\mathbb{R}$, given by

$$(p_1 \circ f)(x_1, x_2, x_3) = v_3x_2 - v_2x_3,$$
$$(p_2 \circ f)(x_1, x_2, x_3) = v_1x_3 - v_3x_1,$$
$$(p_3 \circ f)(x_1, x_2, x_3) = v_2x_1 - v_1x_2.$$

If we can show that each $p_j \circ f$ is continuous on $\mathbb{R}^3$, then it follows from Proposition 1.19 that f is continuous on $\mathbb{R}^3$. We prove that $p_1 \circ f$ is continuous; the proofs for $p_2 \circ f$ and $p_3 \circ f$ are similar.

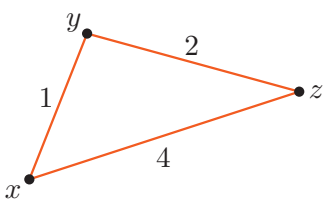
To show that $(p_1 \circ f)(x_1, x_2, x_3) = v_3x_2 - v_2x_3$ is continuous on $\mathbb{R}^3$, we show that the constituent parts are continuous and then combine them using the combination rules.

The functions from $\mathbb{R}^3$ to $\mathbb{R}$ given by $(x_1, x_2, x_3) \mapsto x_2$ and $(x_1, x_2, x_3) \mapsto x_3$ are simply the projection functions p_2 and p_3 on $\mathbb{R}^3$, and these are continuous. Note that we may write

$$(p_1 \circ f)(\mathbf{x}) = v_3p_2(\mathbf{x}) - v_2p_3(\mathbf{x}).$$

Since v_3 and $-v_2$ are fixed real numbers, the Multiple Rule tells us that $(v_3p_2) : \mathbb{R}^3 \to \mathbb{R}$ and $(-v_2p_3) : \mathbb{R}^3 \to \mathbb{R}$ are continuous. Using the Sum Rule, we now conclude that the function $(p_1 \circ f) = (v_3p_2) + (-v_2p_3)$ is continuous.

Solution to Exercise 2.1



The function d is not a metric. It does not satisfy (M3), the Triangle Inequality, since (as shown)

$$d(x, y) + d(y, z) = 1 + 2 = 3 < 4 = d(x, z).$$

Solution to Exercise 2.2

(a) $e_1((0, 0), (1, 0)) = |1 - 0| + |0 - 0| = 1$

(b) $e_1((0, 0), (0, 1)) = |0 - 0| + |1 - 0| = 1$

(c) $e_1((0, 1), (1, 0)) = |1 - 0| + |0 - 1| = 2$

Solution to Exercise 2.3

(a) This is not a metric on $\mathbb{R}$, since there are a and b in $\mathbb{R}$ for which $d(a, b)$ is not real-valued. For example, taking $a = 1$ and $b = 0$ gives $d(a, b) = \sqrt{-1} = i$, an imaginary number. Alternatively, $1 \neq -1$, but $d(1, -1) = 0$ and so (M1) does not hold.

(b) This is not a metric on $\mathbb{R}$, since (M1) does not hold. For example, if $a = b = 1$, then

$$d(a, a) = \sqrt{1 + 1} = \sqrt{2} \neq 0.$$

(c) This is a metric on $\mathbb{R}$. We verify that conditions (M1), (M2) and (M3) hold.

(M1) Note first that $d(a,b) \geq 0$ for each $a, b \in \mathbb{R}$. Also,

$$d(a,a) = \sqrt[3]{|a^3 - a^3|} = 0,$$

and $d(a,b) = 0$ implies that $|b^3 - a^3| = 0$ – that is, $a = b$.

(M2) Note that, for each $a, b \in \mathbb{R}$,

$$d(a,b) = \sqrt[3]{|b^3 - a^3|} = \sqrt[3]{|a^3 - b^3|} = d(b,a).$$

(M3) For each $a, b, c \in \mathbb{R}$, we must show that

$$d(a,c) \leq d(a,b) + d(b,c).$$

Since $d(a,c)$, $d(a,b)$ and $d(b,c)$ are all non-negative, it is enough to show that

$$d(a,c)^3 \leq (d(a,b) + d(b,c))^3.$$

Now the Triangle Inequality for the modulus function gives

$$\begin{aligned} d(a,c)^3 = |c^3 - a^3| &= |c^3 - b^3 + b^3 - a^3| \\ &\leq |c^3 - b^3| + |b^3 - a^3| = d(b,c)^3 + d(a,b)^3. \end{aligned}$$

Since $d(b,c) \geq 0$ and $d(a,b) \geq 0$, we have

$$d(b,c)^3 + d(a,b)^3 \leq (d(b,c) + d(a,b))^3,$$

and so

$$d(a,c)^3 \leq (d(b,c) + d(a,b))^3 = (d(a,b) + d(b,c))^3,$$

as required.

Hence (M1)–(M3) hold and we conclude that d is a metric for $\mathbb{R}$.

Solution to Exercise 2.4

It is certainly the case that $d(\mathbf{a}, \mathbf{b})$ is always non-negative and for each $\mathbf{a} \in \mathbb{R}^2$, $d(\mathbf{a}, \mathbf{a}) = 0$.

Now suppose that $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$ with $d(\mathbf{a}, \mathbf{b}) = 0$. We must show that $\mathbf{a} = \mathbf{b}$.

If $d(\mathbf{a}, \mathbf{b}) = 0$, then $|b_1 - a_1| + d_0(a_2, b_2) = 0$, which, since both $|\cdot|$ and d_0 are always non-negative, implies

$$|b_1 - a_1| = 0 \text{ and } d_0(a_2, b_2) = 0.$$

Hence, since $|\cdot|$ and d_0 are metrics on $\mathbb{R}$ (and so satisfy (M1)), both $a_1 = b_1$ and $a_2 = b_2$. Thus $\mathbf{a} = \mathbf{b}$ as required.

Solution to Exercise 2.5

We need to show that for any choice of $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$, we have $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$. This follows directly from the corresponding properties of $|\cdot|$ and d_0 when considered as metrics on $\mathbb{R}$:

$$\begin{aligned} d(\mathbf{a}, \mathbf{b}) &= |b_1 - a_1| + d_0(a_2, b_2) \\ &= |a_1 - b_1| + d_0(b_2, a_2), \text{ by (M2) for } |\cdot| \text{ and } d_0 \\ &= d(\mathbf{b}, \mathbf{a}), \end{aligned}$$

as required.

Solution to Exercise 2.6

Let $a \in X$ and suppose that $r \geq 0$.

Since $B_{d_0}[a, r] = \{x \in X : d_0(a, x) \leq r\}$ and $d_0(a, x) = 1$ unless $a = x$ (when it is 0), we conclude that

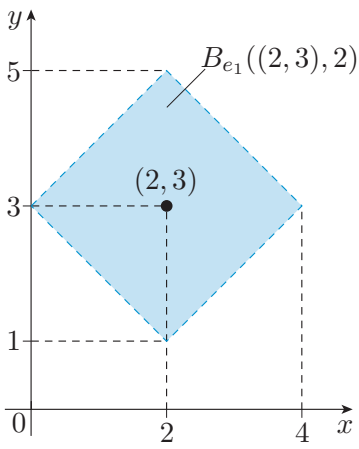
$$B_{d_0}[a, r] = \begin{cases} \{a\}, & \text{if } 0 \leq r < 1, \\ X, & \text{if } r \geq 1. \end{cases}$$

Solution to Exercise 2.7

Let $a \in X$ and suppose that $r > 0$.

Since $S_{d_0}(a, r) = \{x \in X : d_0(a, x) = r\}$ and $d_0(a, x) = 1$ unless $a = x$ (when it is 0), we conclude that

$$S_{d_0}(a, r) = \begin{cases} \{a\}, & \text{if } r = 0, \\ X - \{a\}, & \text{if } r = 1, \\ \emptyset, & \text{otherwise.} \end{cases}$$



Solution to Exercise 2.8

The exercise is to find all $\mathbf{x}$ satisfying

$$e_1((2, 3), \mathbf{x}) = |x_1 - 2| + |x_2 - 3| < 2.$$

The point $(2, 3)$ is the centre of the ball, and its radius is 2. If we imagine the origin of coordinates at $(2, 3)$ and put the vertices of the diamond two units from the centre along the axes, the required figure results (as shown).

Solution to Exercise 2.9

Let $\mathbf{a} = (a_1, a_2) \in \mathbb{R}^2$. If $r = 0$ then by Worked Exercise 2.7, $B_d[\mathbf{a}, r] = \{\mathbf{a}\}$.

Now suppose that $r > 0$. Then

$$\begin{aligned} B_d[\mathbf{a}, r] &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : d(\mathbf{a}, \mathbf{x}) \leq r\} \\ &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : |x_1 - a_1| + d_0(a_2, x_2) \leq r\}. \end{aligned}$$

If $x_2 = a_2$, then $d_0(a_2, x_2) = 0$ and

$$\{\mathbf{x} = (x_1, a_2) \in \mathbb{R}^2 : |x_1 - a_1| \leq r\} = [a_1 - r, a_1 + r] \times \{a_2\}.$$

If $x_2 \neq a_2$, then $d_0(a_2, x_2) = 1$ and

$$\begin{aligned}
&\{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : x_2 \neq a_2 \text{ and } |x_1 - a_1| + d_0(a_2, x_2) \leq r\} \\
&= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : x_2 \neq a_2 \text{ and } |x_1 - a_1| \leq r - 1\} \\
&= \begin{cases} \emptyset, & \text{if } r < 1, \\ [a_1 - (r-1), a_1 + (r-1)] \times (\mathbb{R} - \{a_2\}), & \text{if } r \geq 1. \end{cases}
\end{aligned}$$

Hence

$$\begin{aligned}
&B_d[\mathbf{a}, r] \\
&= \begin{cases} [a_1 - r, a_1 + r] \times \{a_2\}, & \text{if } r < 1, \\ ([a_1 - r, a_1 + r] \times \{a_2\}) & \\ \qquad \cup ([a_1 - (r-1), a_1 + (r-1)] \times \mathbb{R}), & \text{if } r \geq 1. \end{cases}
\end{aligned}$$

Solution to Exercise 3.1

To show that $(\mathbf{a}_k)$ converges to $(1, 2)$ for e_1, we must show that the sequence of real numbers $(e_1(\mathbf{a}_k, (1, 2)))$ is null.

We calculate

$$\begin{aligned}
e_1(\mathbf{a}_k, (1, 2)) &= e_1((1 + \tfrac{1}{2k}, 2 - \tfrac{1}{2k}), (1, 2)) \\
&= \left|1 - (1 + \tfrac{1}{2k})\right| + \left|2 - (2 - \tfrac{1}{2k})\right| = \tfrac{1}{2k} + \tfrac{1}{2k} = \tfrac{1}{k}.
\end{aligned}$$

Since $\left(\frac{1}{k}\right)$ is a basic null sequence, we conclude that $\mathbf{a}_k \to (1, 2)$ with respect to e_1.

Solution to Exercise 3.2

Let (a_k) be a sequence in X and suppose that it converges (in d_0) to $a \in X$.

Then, by the definition of convergence, it must be the case that $(d_0(a_k, a))$ is a real null sequence and so, in particular, there is $N \in \mathbb{N}$ such that for $k > N$, $|d_0(a_k, a)| < 1$.

But $d_0(a_k, a)$ can only equal 1 or 0, so this means that there must be $N \in \mathbb{N}$ such that for $k > N$, $d_0(a_k, a) = 0$. Since d_0 is a metric on X, property (M1) tells us that for $k > N$, $a_k = a$. In other words, the sequence (a_k) is eventually constant, as required.

Solution to Exercise 4.1

We take a general point $a \in X$ and prove that g is $(d, d^{(1)})$-continuous at a. To do this, we take a sequence (x_k) in X with $x_k \xrightarrow{d} a$ as $k \to \infty$ and show that $g(x_k) \xrightarrow{d^{(1)}} g(a)$. This means we must show that

$$d^{(1)}(g(x_k), g(a)) \to 0 \quad \text{as } k \to \infty,$$

and so we find an upper bound for this quantity.

Using the definition of g, and the Reverse Triangle Inequality for e, we find that

$$\begin{aligned}
d^{(1)}(g(x_k), g(a)) &= |g(x_k) - g(a)| \\
&= |e(f(x_k), b) - e(f(a), b)| \\
&\leq e(f(x_k), f(a)). \qquad (2)
\end{aligned}$$

Since f is (d, e)-continuous on X and $x_k \xrightarrow{d} a$ as $k \to \infty$, we have that

$$f(x_k) \xrightarrow{e} f(a).$$

So $(e(f(x_k), f(a)))$ is a null sequence. Hence, applying the Squeeze Rule to (2) shows that $\left(d^{(1)}(g(x_k), g(a))\right)$ is also a null sequence. That is, $(g(x_k))$ converges to $g(a)$ for $d^{(1)}$. Thus g is $(d, d^{(1)})$-continuous at a.

Since the choice of a is arbitrary, g is $(d, d^{(1)})$-continuous on X.

Chapter 15

Metric spaces and continuity 2

Introduction

In this chapter we build upon the work we did in Chapter 14 by increasing our stock of standard examples of metric spaces. We do this in two ways: by finding ways of forming new metric spaces from existing ones and by looking at some new examples of metrics. In particular, we introduce important examples of metric spaces such as sequence and function spaces.

We also take some time for further examination of continuity between metric spaces and introduce an important class of continuous functions, the *Lipschitz functions*.

In Section 4 we look at what it could mean to say that a sequence of functions converges and in Section 5 we use these ideas to define a metric on a space that consists of continuous functions.

Chapter guide

Although we introduce several techniques and ideas in this chapter, they build on the foundations we laid in the last chapter. In many ways this chapter may be viewed as an opportunity to consolidate and deepen your understanding of metrics.

The examples we give are fundamental in many areas of mathematics. In particular, the space of functions that we consider in Section 5 occurs frequently (and will be studied again in Book F) – you should pay particular attention to this section.

1 New metrics from old

In the previous chapter, we saw several examples of metrics and metric spaces. In this section, we look at some methods by which we can extend our collection of 'known' metric spaces. In particular, in Subsection 1.1 we look at how we can transfer a metric 'across' into a different space: this allows us to formalise the idea of subspaces for a metric space. In Subsection 1.3 we look at how to define products of metric spaces and see that there are several ways of doing this.

1.1 Transferring a metric

Our first method for creating new metric spaces is to find a way of using functions to transfer metrics from one set to another.

Suppose that we have a set X on which we wish to define a metric, and that we also have a metric space (W, d). We wish to 'transfer' the metric d to X; that is, we wish to define a metric on X in terms of the metric d on W. At first sight, it may seem that what we need to do is to bring the metric d from W to X by defining a function from W to X. Rather surprisingly, this does not work effectively, but a suitable function *from* X *into* W does – the idea is that we are finding a function that *embeds* X inside W and this embedding then determines the distance between points in X.

So suppose that $f: X \to W$ and consider two points, x and y, in X. Since f maps from X to W, there are points u and v in W for which $u = f(x)$ and $v = f(y)$. We use the distance (in W) between u and v to define the distance between x and y. Since $u = f(x)$ and $v = f(y)$, this leads us to define a distance d_f on X by $d_f(x, y) = d(u, v) = d(f(x), f(y))$, as shown in Figure 1.1: our hope is that this gives us a metric. As we will shortly see, for this to work, we require $f: X \to W$ to be a *one–one* function.

Remember that a function f is *one–one* if $x = y$ whenever $f(x) = f(y)$.

We first give the function d_f a name.

Definition 1.1 *Pull-back*

Let $f: X \to W$ be a function and let d be a metric on W. The **pull-back of d by f** is the function $d_f: X \times X \to \mathbb{R}$ given by

$$d_f(x, y) = d(f(x), f(y)), \text{ for each } x, y \in X.$$

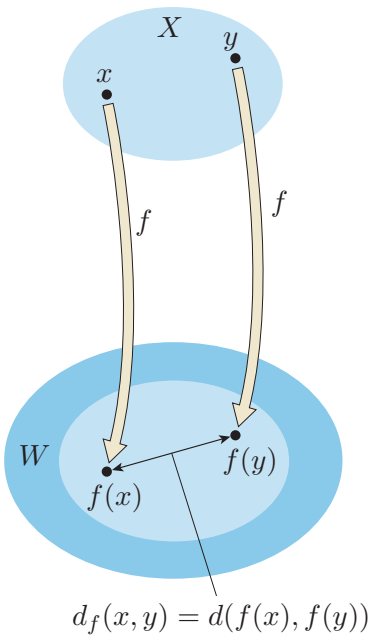


Figure 1.1 Defining $d_f(x, y)$

We now prove that d_f is indeed a metric when f is one–one.

Theorem 1.2

Let $f: X \to W$ be a one–one function and let d be a metric on W. Then the pull-back of d by f, d_f, is a metric on X.

Proof We show that d_f satisfies conditions (M1)–(M3) (Definition 2.1 of Chapter 14).

(M1) The metric d takes only non-negative values. So, for each $x, y \in X$, $d_f(x, y) = d(f(x), f(y)) \geq 0$.

For any $x \in X$, $d_f(x, x) = d(f(x), f(x)) = 0$.

Conversely, if $x, y \in X$ are such that $d_f(x, y) = 0$, then $d(f(x), f(y)) = 0$. Since d is a metric, this implies that $f(x) = f(y)$. But f is one–one, so $x = y$, and d_f satisfies (M1).

This is why f must be one–one.

(M2) For each $x, y \in X$,

$$d_f(y, x) = d(f(y), f(x)) = d(f(x), f(y)) = d_f(x, y),$$

since d is a metric on W. Thus, d_f satisfies (M2).

(M3) Let $x, y, z \in X$. Then, using that d itself satisfies (M3), we find

$$\begin{aligned} d_f(x,z) &= d(f(x), f(z)) \\ &\leq d(f(x), f(y)) + d(f(y), f(z)) \\ &= d_f(x,y) + d_f(y,z). \end{aligned}$$

Thus, d_f satisfies (M3).

Therefore d_f satisfies (M1)–(M3), and so is a metric on X. ■

If f is not one–one, then there are $x, y \in X$ such that $x \neq y$ and $f(x) = f(y)$. For this choice of x and y, we have

$$d_f(x,y) = d(f(x), f(y)) = 0.$$

But this means that d_f fails to satisfy (M1), and so is not a metric. Thus the pull-back of d by f is a metric *if, and only if,* f is one–one.

Worked Exercise 1.3

Show that $d\colon (0,1) \times (0,1) \to \mathbb{R}$ given by $d(x,y) = |y^{-1} - x^{-1}|$ is a metric on the open interval $(0,1)$.

Solution

We could show that d is a metric by verifying directly that it satisfies (M1)–(M3). However, it is easier to observe that if $f\colon (0,1) \to \mathbb{R}$ is given by $f(x) = x^{-1}$, then d is the pull-back of the Euclidean metric $d^{(1)}$ on $\mathbb{R}$ by f, since

$$d(x,y) = |y^{-1} - x^{-1}| = d^{(1)}(f(x), f(y)).$$

Thus, by Theorem 1.2 we need only verify that f is one–one to conclude that d is a metric.

To show that f is one–one, suppose that $x, y \in (0,1)$ and that $f(x) = f(y)$. Then $x^{-1} = y^{-1}$, so that $x = y$. Hence f is one–one, and so d is a metric on $(0,1)$.

Exercise 1.1

Use Theorem 1.2 to show that $d\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defined by

$$d(x,y) = |\tan^{-1} y - \tan^{-1} x|$$

is a metric on $\mathbb{R}$.

Hint: recall that $\tan^{-1}$ is the inverse function of tan, so that $\tan^{-1} x$ is the unique angle $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$ such that $\tan\theta = x$. Its graph is sketched in Figure 1.2.

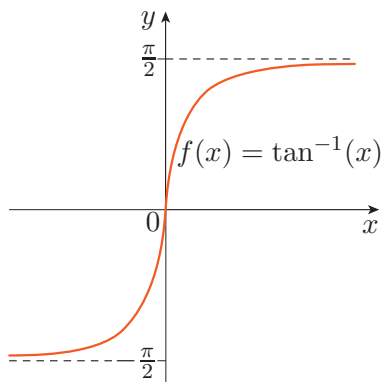


Figure 1.2 Graph of $f(x) = \tan^{-1} x$

Exercise 1.2

Let $d\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined by

$$d(x, y) = |y^2 - x^2|.$$

Determine whether d is a metric on $\mathbb{R}$.

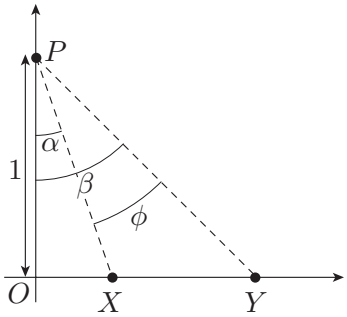


Figure 1.3 Geometric interpretation of the $\tan^{-1}$ metric

The metric defined in Exercise 1.1 is referred to as the **$\tan^{-1}$ metric**, and it has an interesting geometric interpretation.

Consider Figure 1.3. A surveyor at the point P, one unit up the vertical axis, sights the points X and Y (both on the horizontal axis) along the lines PX and PY shown. With P as the vertex, these lines determine an angle ϕ. Let us use ϕ as a measure of the distance d (from the surveyor's perspective) between X and Y.

Let X and Y be at distances x and y respectively from the origin O and assume, without loss of generality, that $y \geq x$: we now find d as a function of x and y. By considering the triangle with vertices P, O and X, we find $\tan\alpha = x$ (since $OP = 1$), so $\alpha = \tan^{-1} x$. Similarly, $\beta = \tan^{-1} y$. Since $y \geq x$, we find that $\phi = \beta - \alpha$ and $d(X, Y) = \phi = \tan^{-1} y - \tan^{-1} x$. In the case $x \geq y$, we would find that $d(X, Y) = \phi = \tan^{-1} x - \tan^{-1} y$. Combining these two cases yields the formula for $d(X, Y)$:

$$d(X, Y) = |\tan^{-1} y - \tan^{-1} x|.$$

Thus, $d(X, Y)$ is the angle between the points X and Y, as observed by the surveyor.

1.2 Metric subspaces

Given a metric on a space X, we often wish to work within a particular subset of our space and use a metric defined on it that 'inherits' properties from the metric on the whole space X. For example, we may be interested in looking at $[0, 1]$ rather than the whole of $\mathbb{R}$ (with the usual Euclidean metric); in such circumstances, we want a metric defined on $[0, 1]$ and not on $\mathbb{R}$ that is consistent with the metric on the whole space.

So suppose we are given a subset A of the underlying set X of a metric space (X, d). We can view A as embedded in X via the function $f\colon A \to X$ given by $f(a) = a$ for each $a \in A$. It is easy to see that this function is one–one and so, from Theorem 1.2, we conclude that the pull-back of d by f, d_f, is a metric on A. But for $a, b \in A$, $d_f(a, b) = d(f(a), f(b)) = d(a, b)$ and so it assigns the right distance between pairs of points in A. We record this in the following definition.

Definition 1.4 *Induced metric*

Let (X, d) be a metric space and let $A \subseteq X$.

The function $d_A \colon A \times A \to \mathbb{R}$ given by

$$d_A(a, b) = d(a, b), \quad \text{for each } a, b \in A,$$

is a metric on A and is called the metric on A **induced by** d, often referred to as the **induced** (or **subspace**) **metric** on A.

The metric space (A, d_A) is called a **metric subspace** of (X, d), and often referred to as the **induced subspace**.

Note that if it is clear from the context, then we may suppress mention of the subset A. For example, we often talk of the Euclidean metric for $[0, 1]$ as $d^{(1)}$ rather than the more cumbersome $d^{(1)}_{[0,1]}$.

Since the metric on a subspace comes from the metric on the original space, one would expect that many of the properties of the subspace will be similar to those of the original space.

Theorem 1.5

Let (X, d) be a metric space and suppose that $A \subseteq X$. Let d_A denote the metric on A induced by d. Then a sequence $(a_n)_{n\in\mathbb{N}}$ in A is d_A-convergent to a if, and only if, $a \in A$ and the sequence (a_n) is d-convergent to a (when viewed as a sequence in X).

Proof Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in A. Then it is also a sequence in X.

Suppose that $a_n \xrightarrow{d_A} a$ as $n \to \infty$. Then, by the definition of convergence, $a \in A$ and $d_A(a_n, a) \to 0$ as $n \to \infty$.

But, by definition, for each $n \in \mathbb{N}$, $d_A(a_n, a) = d(a_n, a)$ and also $a \in A \subseteq X$. Hence $a_n \xrightarrow{d} a$ as $n \to \infty$.

Conversely, suppose that $a \in A$ and the sequence (a_n) d-converges to a as $n \to \infty$.

Then $(d(a_n, a))_{n\in\mathbb{N}}$ is a (real) null sequence and, since $a \in A$ and $a_n \in A$ for each n,

$$d(a_n, a) = d_A(a_n, a) \text{ for each } n \in \mathbb{N}.$$

Hence $(d_A(a_n, a))$ is a (real) null sequence and we conclude that $a_n \xrightarrow{d_A} a$ as $n \to \infty$. ■

Theorem 1.6

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$ is a function. Let $A \subseteq X$ and let d_A be the metric induced on A by d.

If $a \in A$ and f is (d, e)-continuous at a, then $f|_A$ is (d_A, e)-continuous at a.

(Here the **restriction of f to A**, $f|_A\colon A \to Y$, is given by $f|_A(a) = f(a)$ for each $a \in A$.)

Exercise 1.3

Prove Theorem 1.6.

The idea of an induced metric gives us an easy way to define a metric on common geometric surfaces in $\mathbb{R}^3$, such as the unit sphere with centre $\mathbf{0}$:

$$S = S_{d^{(3)}}(\mathbf{0}, 1) = \{\mathbf{x} \in \mathbb{R}^3 : d^{(3)}(\mathbf{0}, \mathbf{x}) = 1\}.$$

We simply endow S with the metric induced by the Euclidean metric $d^{(3)}$.

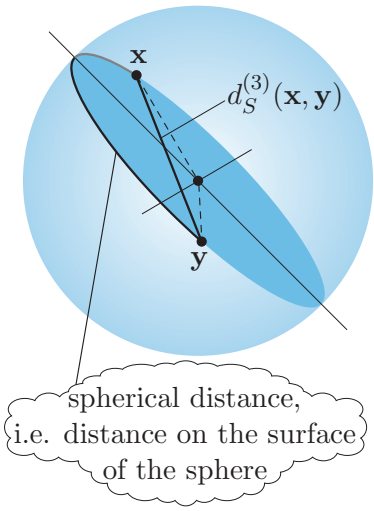


Figure 1.4 Geometric interpretation of $d_S^{(3)}(x, y)$

Exercise 1.4

Show that $(1, 0, 0)$ and $(0, 0, 1)$ are in S, and calculate $d_S^{(3)}((1, 0, 0), (0, 0, 1))$.

Although endowing S with the metric induced by $d^{(3)}$ gives us a metric on the unit sphere, it does not measure distance *on* the sphere in the way that we do on the Earth, as Figure 1.4 illustrates. It is possible to define a metric that does this, but it is a little messy and we omit it.

Exercise 1.5

Let S be the shape given in Figure 1.5, i.e. S consists of three sides of a unit square:

$$S = \{(0, y) : 0 \leq y \leq 1\} \cup \{(x, 0) : 0 \leq x \leq 1\} \cup \{(1, y) : 0 \leq y \leq 1\}.$$

(You could think of S as being a piece of string of length 3 laid around the sides of a square of side 1.)

Let $f\colon S \to \mathbb{R}$ be given by

$$f(x, y) = \begin{cases} 1 - y, & \text{if } x = 0 \text{ and } 0 < y \leq 1, \\ x + 1, & \text{if } y = 0, \\ y + 2, & \text{if } x = 1 \text{ and } 0 < y \leq 1. \end{cases}$$

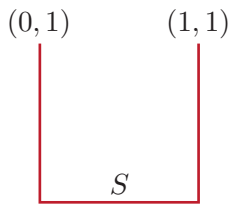


Figure 1.5 S consists of three sides of a unit square: the left side, the base side and the right side

Define a distance function $d\colon S \times S \to \mathbb{R}$ by $d(\mathbf{a}, \mathbf{b}) = |f(\mathbf{a}) - f(\mathbf{b})|$.

(a) Show that f is a one–one function.

(b) Show that d is a metric for S.

(c) For $\mathbf{a} = (0, 1)$ and $\mathbf{b} = (1, 1)$, find $d(\mathbf{a}, \mathbf{b})$ and $d_S^{(2)}(\mathbf{a}, \mathbf{b})$, where $d_S^{(2)}$ is the metric on S induced by the Euclidean metric on $\mathbb{R}^2$.

1.3 Product metrics

We have now seen how to transfer a metric from one set to another (or to a subset of itself) by a one–one function. However, suppose instead that we have a set X defined as the *product* $X_1 \times X_2$ of two sets, each with a metric defined on it. Can we construct a metric on the product set X from the metrics on X_1 and X_2?

$X_1 \times X_2 = \{(x_1, x_2) : x_1 \in X_1 \text{ and } x_2 \in X_2\}$

You have already come across three examples where this can be done, all involving $X = \mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, where $X_1 = X_2 = \mathbb{R}$ with the Euclidean metric $d^{(1)}$. The first is the Euclidean metric $d^{(2)}$ on $\mathbb{R}^2$, which can be written in terms of the Euclidean metric $d^{(1)}$ on $\mathbb{R}$ as

$$\begin{aligned} d^{(2)}(\mathbf{x}, \mathbf{y}) &= \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} \\ &= \sqrt{(d^{(1)}(x_1, y_1))^2 + (d^{(1)}(x_2, y_2))^2}. \end{aligned}$$

The second is the taxicab metric on $\mathbb{R}^2$, which can be written in terms of the Euclidean metric $d^{(1)}$ on $\mathbb{R}$ as

$$e_1(\mathbf{x}, \mathbf{y}) = d^{(1)}(x_1, y_1) + d^{(1)}(x_2, y_2).$$

And the third is the 'mixed' metric d on $\mathbb{R}^2$:

$$d(\mathbf{x}, \mathbf{y}) = d^{(1)}(x_1, y_1) + d_0(x_2, y_2) = |y_1 - x_1| + d_0(x_2, y_2),$$

where d_0 denotes the discrete metric for $\mathbb{R}$.

In fact, if (X_1, d_1) and (X_2, d_2) are metric spaces, then there are many metrics that we can construct on the product set $X = X_1 \times X_2$ from the metrics d_1 and d_2. Such metrics are referred to as **product metrics**. We look in detail at three of them.

Theorem 1.7

Let (X_1, d_1) and (X_2, d_2) be metric spaces, and let $X = X_1 \times X_2$. The functions e_1, e_2 and e_∞ from $X \times X$ to $\mathbb{R}$, given by

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= d_1(x_1, y_1) + d_2(x_2, y_2), \\ e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{(d_1(x_1, y_1))^2 + (d_2(x_2, y_2))^2}, \\ e_\infty(\mathbf{x}, \mathbf{y}) &= \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}, \end{aligned}$$

for each $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2) \in X$, are metrics on X.

Remarks

1. The subscripts 1 and 2 indicate that e_1 and e_2 are formed by using first and second powers respectively. More generally, if we consider an arbitrary power $p \geq 1$ and define

$$e_p(\mathbf{x}, \mathbf{y}) = (d_1(x_1, y_1)^p + d_2(x_2, y_2)^p)^{1/p},$$

We will not prove this.

then e_p is also a metric on X. Moreover, if we take the limit as $p \to \infty$, then we find that $e_p(\mathbf{x}, \mathbf{y}) \to e_\infty(\mathbf{x}, \mathbf{y})$.

2. In the proof that e_∞ is a metric, we use the following fact:

$$\text{if } a, b, c, d \in \mathbb{R}, \text{ then } \max\{a+b, c+d\} \leq \max\{a, c\} + \max\{b, d\}. \quad (1)$$

This follows easily by considering $a + b$ and $c + d$ separately, and by noting that each of these is at most $\max\{a, c\} + \max\{b, d\}$.

3. The 'mixed' metric that we defined for $\mathbb{R}^2$ in Chapter 14 is formed by using the e_1 construction for the metric spaces $(\mathbb{R}, d^{(1)})$ and $(\mathbb{R}, d_0)$.

You are asked to provide this proof in Exercise 1.6.

Proof The proof that e_1 is a metric is very similar to the proof that the taxicab metric on $\mathbb{R}^2$ is a metric.

The proof that e_2 is a metric is very similar to the proof (given in Chapter 13) that the Euclidean metric for $\mathbb{R}^2$, $d^{(2)}$, satisfies conditions (M1)–(M3).

Here we content ourselves with verifying that e_∞ satisfies conditions (M1)–(M3).

(M1) For each $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$, both $d_1(x_1, y_1)$ and $d_2(x_2, y_2)$ are non-negative, since d_1 and d_2 are metrics, and hence so is their maximum: thus $e_\infty(\mathbf{x}, \mathbf{y}) \geq 0$.

For each $\mathbf{x} \in X_1 \times X_2$,

$$e_\infty(\mathbf{x}, \mathbf{x}) = \max\{d_1(x_1, x_1),\, d_2(x_2, x_2)\} = \max\{0, 0\} = 0,$$

since d_1 and d_2 are metrics.

Conversely, suppose $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$ are such that $e_\infty(\mathbf{x}, \mathbf{y}) = 0$. Then

$$0 = \max\{d_1(x_1, y_1),\, d_2(x_2, y_2)\}.$$

Since d_1 and d_2 are metrics, $d_1(x_1, y_1) \geq 0$ and $d_2(x_2, y_2) \geq 0$, so this equation implies that $d_1(x_1, y_1) = 0$ and $d_2(x_2, y_2) = 0$. Thus, since d_1 and d_2 are metrics, $x_1 = y_1$ and $x_2 = y_2$; that is, $\mathbf{x} = \mathbf{y}$. We conclude that e_∞ satisfies (M1).

(M2) Let $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$. Using property (M2) for the metrics d_1 and d_2 respectively, we have

$$\begin{aligned} e_\infty(\mathbf{y}, \mathbf{x}) &= \max\{d_1(y_1, x_1),\, d_2(y_2, x_2)\} \\ &= \max\{d_1(x_1, y_1),\, d_2(x_2, y_2)\} = e_\infty(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus e_∞ satisfies (M2).

(M3) Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X_1 \times X_2$. Then, using property (M3) for the metrics d_1 and d_2,

$$\begin{aligned} e_\infty(\mathbf{x}, \mathbf{z}) &= \max\{d_1(x_1, z_1),\, d_2(x_2, z_2)\} \\ &\leq \max\{d_1(x_1, y_1) + d_1(y_1, z_1),\, d_2(x_2, y_2) + d_2(y_2, z_2)\} \\ &\leq \max\{d_1(x_1, y_1),\, d_2(x_2, y_2)\} + \max\{d_1(y_1, z_1),\, d_2(y_2, z_2)\} \\ &= e_\infty(\mathbf{x}, \mathbf{y}) + e_\infty(\mathbf{y}, \mathbf{z}). \end{aligned}$$

This step is justified by equation (1).

Thus e_∞ satisfies (M3).

Therefore e_∞ is a metric on $X_1 \times X_2$. ■

Exercise 1.6

Let (X_1, d_1) and (X_2, d_2) be metric spaces. Prove that $(X_1 \times X_2, e_1)$ is a metric space.

Product metrics on the plane

When $(X_1, d_1) = (X_2, d_2) = (\mathbb{R}, d^{(1)})$, our methods for constructing metrics on product spaces give us three different ways of defining a metric on $\mathbb{R}^2$:

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= d^{(1)}(x_1, y_1) + d^{(1)}(x_2, y_2) = |y_1 - x_1| + |y_2 - x_2|, \\ e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{d^{(1)}(x_1, y_1)^2 + d^{(1)}(x_2, y_2)^2} = \sqrt{|y_1 - x_1|^2 + |y_2 - x_2|^2}, \\ e_\infty(\mathbf{x}, \mathbf{y}) &= \max\{d^{(1)}(x_1, y_1), d^{(1)}(x_2, y_2)\} = \max\{|y_1 - x_1|, |y_2 - x_2|\}. \end{aligned}$$

The metrics e_1 and e_2 are the familiar taxicab and Euclidean metrics on the plane. The metric e_∞ is known as the **max metric** on the plane.

Exercise 1.7

Show that e_1, e_2 and e_∞ are distinct metrics on the plane by calculating $e_1(\mathbf{0}, (2, 1))$, $e_2(\mathbf{0}, (2, 1))$ and $e_\infty(\mathbf{0}, (2, 1))$.

We can obtain an idea of the difference in behaviour between these metrics on the plane by sketching their unit open balls centred at the origin, as shown in Figure 1.6. In Worked Exercise 2.9 of Chapter 14, we saw that for e_1, the taxicab metric, the unit open ball centred at the origin consists of all the points inside a diamond centred at the origin. Since for the plane $e_2 = d^{(2)}$, the Euclidean distance on the plane, we see that for e_2 the unit open ball centred at the origin is an open disc of unit radius centred at the origin. For e_∞, the unit open ball centred at the origin consists of all the points inside the square of side length 2 centred at the origin. In all three cases, since we are considering *open* balls, the boundary points are excluded.

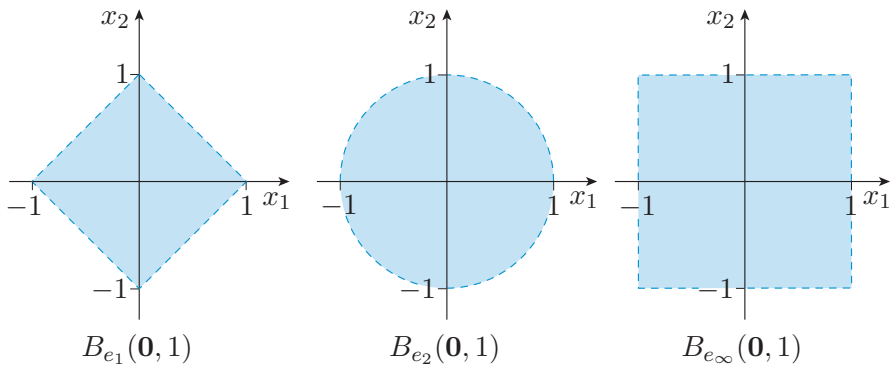


Figure 1.6 Unit open balls centred at the origin for three metrics

Exercise 1.8

(a) Justify the sketch of the unit open ball centred at the origin for e_∞.

(b) What are the unit closed balls and spheres centred at the origin for e_1, e_2 and e_∞?

1.4 n-fold products

Suppose that we are interested in the product of three metric spaces. In principle, this is no more difficult than the product of two: we take the product of the first two, which is a metric space; then we take the product of this new metric space with the third. In doing this, we may use e_1, e_2 or e_∞, or any other product metric that we have defined.

For example, suppose that (X, d) is a metric space; then there are many ways we can define a product metric e on $X^3 = X \times X \times X$. We could take

$$e((a_1, a_2, a_3), (b_1, b_2, b_3)) = d(a_1, b_1) + d(a_2, b_2) + d(a_3, b_3)$$

or

$$e((a_1, a_2, a_3), (b_1, b_2, b_3)) = \sqrt{d(a_1, b_1)^2 + (d(a_2, b_2) + d(a_3, b_3))^2}$$

or many other variants. In all cases, by keeping a clear head and using our results concerning products of pairs of spaces, we can show that these distances e all define metrics on X^3.

This approach can be continued, enabling us to obtain products of n metric spaces, for any given $n \in \mathbb{N}$. We do not go into any great detail here but do take the opportunity to introduce a new example of a metric space in light of this idea.

Example 1.8

Let A be a finite set, called the **set of symbols**, and let d_0 be the discrete metric for A.

Fix $n \in \mathbb{N}$ and let $X = \underbrace{A \times \cdots \times A}_{n} = A^n$, **the words of length n over the set of symbols A.**

Define the Hamming distance $d_H : X \times X \to \mathbb{R}$ for $\mathbf{a}, \mathbf{b} \in X$ by

$$d_H(\mathbf{a}, \mathbf{b}) = \text{the number of places at which } \mathbf{a} \text{ and } \mathbf{b} \text{ differ.}$$

Richard Hamming (1915–1998) was a North American mathematician renowned for his work on error-correcting codes.

It is not too difficult to see that the Hamming distance between pairs of words in X is the product metric e obtained by using e_1, so that for each pair of words $\mathbf{a} = (a_1, \ldots, a_n), \mathbf{b} = (b_1, \ldots, b_n) \in X$,

$$e(\mathbf{a}, \mathbf{b}) = d_0(a_1, b_1) + d_0(a_2, b_2) + \cdots + d_0(a_n, b_n).$$

Hence d_H defines a metric on X.

Remarks

1. It is quite common when considering Hamming distance on A^n to write words as strings $a_1 a_2 \cdots a_n$ rather than as vectors $(a_1, a_2, \ldots, a_n)$.
2. A particularly useful case is obtained when $A = \{0, 1\}$ and so A^n consists of all possible binary strings of length n. In this case, the Hamming distance measures in how many places two binary strings of the same length differ – this has applications in the theory of error-correcting codes.

Balls for the Hamming distance

We look briefly at the balls for the Hamming distance on A^n when $A = \{0, 1\}$.

Worked Exercise 1.9

For $n = 2$, find $B_{d_H}((0,0), r)$ for $r > 0$.

Solution

When $n = 2$, $A^n = A^2 = \{(0,0), (1,0), (0,1), (1,1)\}$ and this set of binary strings can be represented as the vertices of a square. Adjacent vertices are distance 1 apart and non-adjacent vertices are distance 2 apart (see Figure 1.7).

Thus

$$B_{d_H}((0,0), r) = \begin{cases} \{(0,0)\}, & \text{if } 0 < r \leq 1, \\ \{(0,0), (1,0), (0,1)\}, & \text{if } 1 < r \leq 2, \\ \{(0,0), (1,0), (0,1), (1,1)\} \, (= A^2), & \text{if } 2 < r. \end{cases}$$

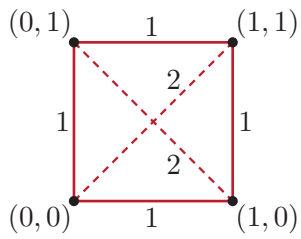


Figure 1.7 Representation of A^2

Exercise 1.9

For $n = 3$, find $B_{d_H}((0,0,0), r)$ for $r > 0$.

Hamming distance can be used as a basis for defining error-correcting codes, as follows. Suppose that we have a string of zeros and ones to transmit. We encode zero as the string 000 and one as the string 111, so the message 010 is transmitted as 000111000. The receiver splits the received message into 'blocks' of length three, and decodes the blocks as 0 if they differ from 000 in no more than one place and 1 if they differ from 111 in no more than one place. It is quite easy to see that following this method, provided no more than one error occurs in each block of length three, the message can still be accurately decoded. Of course, this method of encoding triples the length of each message and there are other, rather more efficient approaches.

2 The Cantor metric

In this section, we give an example of a metric space that is a *sequence space*. Sequence spaces have applications in many areas of mathematics and are often used as a model for understanding the behaviour and structure of more complicated objects, such as *function spaces*.

We discuss an example of a function space later.

Definition 2.1 *Cantor space*

The **Cantor space C** consists of all *infinite* sequences of zeros and ones:

$$\mathbf{C} = \{(a_n) : a_n \in \{0,1\} \text{ for each } n \in \mathbb{N}\}.$$

You can learn more about Georg Cantor (1845–1918) in the *History Reader*.

Remarks

This is a natural extension of the notation for vectors and suggests that one possible way of thinking about points in **C** is to consider them to be infinite vectors.

1. For ease of writing, we denote points in **C** by bold letters: so $\mathbf{x} = (x_n)$ denotes the point $(x_1, x_2, x_3, \ldots)$.
2. Two points in **C** are the same if, and only if, all their terms agree; that is, $\mathbf{x} = \mathbf{y}$ if, and only if, $x_n = y_n$ for each $n \in \mathbb{N}$.
3. There is a useful way of visualising Cantor space in terms of an infinite binary tree, as shown in Figure 2.1.

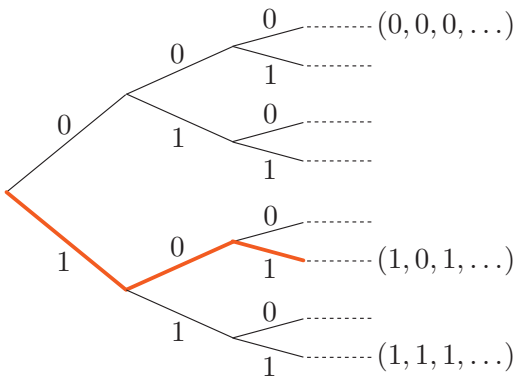


Figure 2.1 An infinite binary tree representing Cantor space

Examples of points in $\mathbf{C}$ are infinite sequences such as

$$(1,1,1,1,\ldots,1,\ldots), \quad (0,0,0,0,\ldots,0,\ldots), \quad (1,0,1,0,\ldots,1,0,\ldots).$$

We have highlighted the start of the path in the tree that represents $(1,0,1,0,\ldots,1,0,\ldots)$.

Suppose we are given two points $\mathbf{x} = (x_n)$ and $\mathbf{y} = (y_n)$ in $\mathbf{C}$. How can we measure the distance between them? We already have one metric that could be used: d_0, the discrete metric. This *would* give us a notion of distance on $\mathbf{C}$ but it would be one that reflected nothing of the structure of $\mathbf{C}$ – we desire a notion of distance that reflects something of the definition of elements of $\mathbf{C}$. One idea would be to follow the concept behind the definition of Hamming distance and count the number of places (or terms) where $\mathbf{x}$ and $\mathbf{y}$ differ. Unfortunately, $\mathbf{x}$ and $\mathbf{y}$ may differ at *infinitely* many terms; thus this definition of distance would have to take infinite values and so would certainly not define a metric.

Consider $(0,0,0,0,\ldots)$ and $(1,1,1,1,\ldots)$, for example.

There are various ways to resolve this problem. The approach we take is to find the first term at which the two sequences differ, and use this as a way of defining the distance between them. Intuitively, such a definition of distance should give smaller values the further 'down' the sequences we must go before we find a term at which the sequences differ.

With this in mind, consider the following four sequences:

$$\mathbf{a} = (0,0,0,0,0,\ldots); \quad \mathbf{b} = (1,1,0,0,0,\ldots);$$
$$\mathbf{c} = (1,0,0,0,0,\ldots); \quad \mathbf{d} = (1,0,1,1,1,\ldots).$$

The sequence $\mathbf{a}$ first differs from $\mathbf{b}$ at the first term; we could say that they are a distance $2^{-1} = \frac{1}{2}$ apart. Next, $\mathbf{b}$ and $\mathbf{c}$ first differ at the second term; we could say that they are a distance $2^{-2} = \frac{1}{4}$ apart. Finally, $\mathbf{c}$ and $\mathbf{d}$ first differ at the third term; we could say that they are a distance $2^{-3} = \frac{1}{8}$ apart. In this way, we get smaller distances the further along the two sequences we have to go before they differ and, of course, if the terms of the two sequences are identical, $x_n = y_n$ for each n, then $\mathbf{x} = \mathbf{y}$.

This leads to the following definition of distance in $\mathbf{C}$.

Definition 2.2 *Cantor distance*

Let $\mathbf{x}, \mathbf{y} \in \mathbf{C}$. The **Cantor distance** $d_{\mathbf{C}} : \mathbf{C} \times \mathbf{C} \to \mathbb{R}$ between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$d_{\mathbf{C}}(\mathbf{x}, \mathbf{y}) = \begin{cases} 0, & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-n}, & \text{if } \mathbf{x} \text{ and } \mathbf{y} \text{ first differ at their } n\text{th term.} \end{cases}$$

Note that the Cantor distance between *any* two points of $\mathbf{C}$ is at most $\frac{1}{2}$.

Exercise 2.1

Find $d_{\mathbf{C}}(\mathbf{a}, \mathbf{e})$, $d_{\mathbf{C}}(\mathbf{b}, \mathbf{e})$, $d_{\mathbf{C}}(\mathbf{c}, \mathbf{e})$ and $d_{\mathbf{C}}(\mathbf{d}, \mathbf{e})$ when

$$\mathbf{a} = (1, 0, 0, 0, 0, 0, 0, \ldots), \quad \mathbf{b} = (1, 1, 0, 0, 0, 0, 0, \ldots),$$
$$\mathbf{c} = (1, 1, 1, 0, 0, 0, 0, \ldots), \quad \mathbf{d} = (1, 1, 1, 1, 0, 0, 0, \ldots),$$
$$\mathbf{e} = (1, 1, 1, 1, 1, 1, 1, \ldots).$$

We now show that $d_{\mathbf{C}}$ is a metric on $\mathbf{C}$.

Theorem 2.3

The Cantor distance is a metric on $\mathbf{C}$.

See Chapter 14, Definition 2.1.

Proof We verify that $d_{\mathbf{C}}$ satisfies conditions (M1)–(M3).

(M1) For each $\mathbf{a}, \mathbf{b} \in \mathbf{C}$, we have by definition that $d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) \geq 0$. Moreover $d_{\mathbf{C}}(\mathbf{a}, \mathbf{a}) = 0$.

If $d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 0$ then, by definition, $\mathbf{a} = \mathbf{b}$ and so $d_{\mathbf{C}}$ satisfies (M1).

(M2) Let $\mathbf{a}, \mathbf{b} \in \mathbf{C}$. If $\mathbf{a} = \mathbf{b}$ then, by definition,

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = d_{\mathbf{C}}(\mathbf{b}, \mathbf{a}) = 0.$$

If instead $\mathbf{a} \neq \mathbf{b}$ and $\mathbf{a}$ and $\mathbf{b}$ first differ at the nth term, then $\mathbf{b}$ and $\mathbf{a}$ also first differ at the nth term. Hence,

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = d_{\mathbf{C}}(\mathbf{b}, \mathbf{a}) = 2^{-n}.$$

Thus (M2) holds.

(M3) Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{C}$. We consider four possibilities: (i) $\mathbf{a} = \mathbf{b}$, (ii) $\mathbf{b} = \mathbf{c}$, (iii) $\mathbf{a} = \mathbf{c}$, and (iv) $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$ are all distinct from each other.

(i) If $\mathbf{a} = \mathbf{b}$, so that $d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 0$, then

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c})$$

and so (M3) holds. (In fact $d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}) = d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c})$ here.)

(ii) A similar argument applies if $\mathbf{b} = \mathbf{c}$.

(iii) If $\mathbf{a} = \mathbf{c}$ then $d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = 0$, and once again (M3) holds.

(iv) It only remains to consider the case when **a**, **b** and **c** are all distinct. Suppose that **a** first differs from **b** at the mth term and **a** first differs from **c** at the nth term. There are two possible cases: $m \leq n$ and $m > n$.

If $m \leq n$ (see Figure 2.2), then

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = 2^{-n} \leq 2^{-m} = d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}), \text{ by (M1)}.$$

If $m > n$ (see Figure 2.3), then **b** and **c** must differ first at the nth term (since **a** agrees with both of them for the first $n-1$ terms and agrees with **b** but not with **c** at the nth term). Thus,

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}).$$

It follows that

$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{c}) = 2^{-n} = d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) + d_{\mathbf{C}}(\mathbf{b}, \mathbf{c}), \text{ by (M1)}.$$

Thus, in both cases, (M3) holds.

Since $d_{\mathbf{C}}$ satisfies (M1)–(M3), it is a metric on $\mathbf{C}$. ■

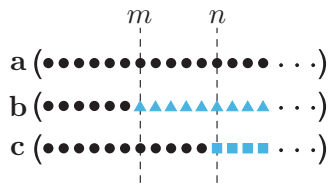


Figure 2.2 Case where $m \leq n$

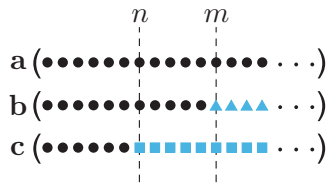


Figure 2.3 Case where $m > n$

Since the Cantor distance is a metric, it is often referred to as the **Cantor metric**.

There are many other examples of sequence spaces in mathematics. Some of the most useful involve sequences of real numbers with specific properties. The method we have given here serves as a model for defining a metric in these more sophisticated settings. In Exercise 2.2, we ask you to show that another sequence space is a metric space.

Exercise 2.2

We say that a sequence of real numbers (a_n) is *eventually zero* if there is an $N \in \mathbb{N}$ such that $a_n = 0$ for each $n > N$. Let ℓ_0 denote the set consisting of all such sequences; that is,

This is a special case of the 'eventually constant' sequence defined in Chapter 14.

$$\ell_0 = \{(a_n) : a_n \in \mathbb{R} \text{ and } (a_n) \text{ is eventually zero}\}.$$

(Notice that different elements of ℓ_0 may have different numbers of non-zero terms.)

Define $d\colon \ell_0 \times \ell_0 \to \mathbb{R}$ by

$$d((a_n), (b_n)) = \sum_{n=1}^{\infty} |b_n - a_n|.$$

This sum is always finite since only finitely many of the terms a_n and b_n are non-zero.

Prove that d defines a metric on ℓ_0.

Balls for the Cantor metric

We now describe open and closed balls for the Cantor metric.

Worked Exercise 2.4

Let $\mathbf{a} = (a_1, a_2, \ldots) \in \mathbf{C}$ and $n \in \mathbb{N}$. Show that for the Cantor metric d,

$$B_d(\mathbf{a}, 2^{-n}) = \{(a_1, a_2, a_3, \ldots, a_n, c_1, c_2, c_3, \ldots) : (c_i) \in \mathbf{C}\}.$$

In other words, the open ball around $\mathbf{a}$ of radius 2^{-n} consists of all infinite sequences of zeros and ones for which the first n terms agree with the first n terms of $\mathbf{a}$ (as shown in Figure 2.4).

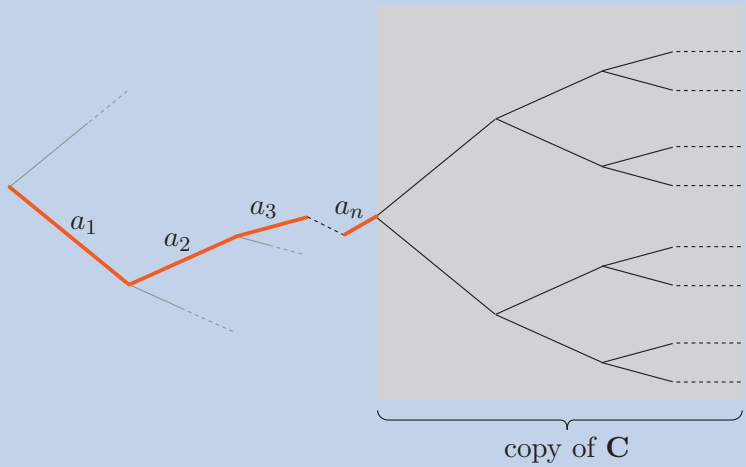


Figure 2.4 The open ball around $\mathbf{a}$ of radius 2^{-n}

Solution

Let $\mathbf{a} = (a_1, a_2, \ldots) \in \mathbf{C}$ and $n \in \mathbb{N}$, and recall that by definition

$$B_d(\mathbf{a}, 2^{-n}) = \{\mathbf{b} \in \mathbf{C} : d(\mathbf{a}, \mathbf{b}) < 2^{-n}\}.$$

Hence if $\mathbf{b} \in \mathbf{C}$ and there is an $i \in \{1, 2, \ldots, n\}$ for which $a_i \neq b_i$, then

$$d(\mathbf{a}, \mathbf{b}) \geq 2^{-i} \geq 2^{-n}$$

and so such a $\mathbf{b}$ is not in $B_d(\mathbf{a}, 2^{-n})$.

Conversely, if $\mathbf{b} \in \mathbf{C}$ agrees with $\mathbf{a}$ for the first n terms, then $\mathbf{b} = (a_1, a_2, \ldots, a_n, b_{n+1}, b_{n+2}, \ldots)$ and $\mathbf{b}$ cannot differ from $\mathbf{a}$ before the $(n+1)$th term. In which case, $d(\mathbf{a}, \mathbf{b}) \leq 2^{-n-1} < 2^{-n}$, implying that $\mathbf{b} \in B_d(\mathbf{a}, 2^{-n})$.

We conclude that

$$B_d(\mathbf{a}, 2^{-n}) = \{(a_1, a_2, a_3, \ldots, a_n, c_1, c_2, c_3, \ldots) : (c_i) \in \mathbf{C}\},$$

as required.

Exercise 2.3

Let $\mathbf{a} = (a_1, a_2, \ldots) \in \mathbf{C}$ and $n \in \mathbb{N}$. Show that for the Cantor metric d,

$$B_d[\mathbf{a}, 2^{-n}] = \begin{cases} \mathbf{C}, & \text{if } n = 1, \\ \{(a_1, a_2, a_3, \ldots, a_{n-1}, c_1, c_2, c_3, \ldots) : (c_i) \in \mathbf{C}\}, & \text{if } n \geq 2. \end{cases}$$

In other words, the closed ball around $\mathbf{a}$ of radius 2^{-n} consists of all infinite sequences of zeros and ones for which the first $n-1$ terms agree with the first $n-1$ terms of $\mathbf{a}$.

Continuity and the Cantor metric

The Cantor space occurs in many different areas of mathematics and there are many examples of continuous functions defined upon it. One simple example is the following.

Definition 2.5 *Shift map*

The **shift map (on C)** is the map $\sigma\colon \mathbf{C} \to \mathbf{C}$ given by

$$\sigma((x_1, x_2, x_3, x_4, \ldots)) = (x_2, x_3, x_4, x_5, \ldots).$$

Thus the shift map acts by deleting the first term of the given sequence. For example,

$$\sigma((0,1,1,1,1,1,\ldots)) = (1,1,1,1,1,\ldots),$$
$$\sigma((0,1,0,1,0,1,\ldots)) = (1,0,1,0,1,\ldots),$$
$$\sigma((1,1,1,1,1,1,\ldots)) = (1,1,1,1,1,\ldots).$$

The shift map 'shifts' the sequence by one term. Such maps play an important role in the theory of dynamical systems.

Worked Exercise 2.6

Show that the shift map is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous on $\mathbf{C}$.

Solution

Let $\mathbf{a} \in \mathbf{C}$ and let $(\mathbf{x}_k)$ be a sequence in $\mathbf{C}$ that converges to $\mathbf{a}$ for $d_{\mathbf{C}}$. We must show that $(\sigma(\mathbf{x}_k))$ converges to $\sigma(\mathbf{a})$ for $d_{\mathbf{C}}$.

To do this, we find an upper bound for $d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a}))$ in terms of $d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a})$. We consider three cases:

1. If $\mathbf{x}_k = \mathbf{a}$, then $d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a}) = 0$ and $d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a})) = 0$.
2. If $\mathbf{x}_k \neq \mathbf{a}$, and $\mathbf{x}_k$ and $\mathbf{a}$ differ at the first term, then $d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a}) = \frac{1}{2}$ and $d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a})) \leq \frac{1}{2}$, since the Cantor distance between two points of $\mathbf{C}$ cannot exceed $\frac{1}{2}$. Thus,
$$d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a})) \leq \tfrac{1}{2} = d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a}).$$

Note that $(\mathbf{x}_k)$ is a sequence of 'points' in the Cantor space $\mathbf{C}$. So for a fixed value of k,
$$\mathbf{x}_k = (x_{k,1}, x_{k,2}, x_{k,3}, \ldots)$$
is a sequence of zeros and ones.

3. If $\mathbf{x}_k \neq \mathbf{a}$, but their first terms are equal, then they must *first* differ at some later term. Suppose that they first differ at the Nth term, where $N \geq 2$. Then $\sigma(\mathbf{x}_k)$ and $\sigma(\mathbf{a})$ first differ at the $(N-1)$th term, and so

$$d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a})) = 2^{-(N-1)} = 2 \times 2^{-N} = 2d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a}).$$

In all three cases,

$$0 \leq d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a})) \leq 2d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a}).$$

Since $(\mathbf{x}_k)$ converges to $\mathbf{a}$ for $d_{\mathbf{C}}$, we have that $d_{\mathbf{C}}(\mathbf{x}_k, \mathbf{a}) \to 0$ as $k \to \infty$. Hence, applying the Squeeze Rule to the inequality above shows that

$$d_{\mathbf{C}}(\sigma(\mathbf{x}_k), \sigma(\mathbf{a})) \to 0 \text{ as } k \to \infty.$$

It follows that $(\sigma(\mathbf{x}_k))$ converges to $\sigma(\mathbf{a})$ for $d_{\mathbf{C}}$, and hence σ is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous at $\mathbf{a}$. Since $\mathbf{a}$ is an arbitrary point of $\mathbf{C}$, we conclude that σ is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous on $\mathbf{C}$.

3 Equivalent metrics

In this section we investigate some conditions under which different metrics on the same underlying set can give rise to the same continuous functions.

3.1 Lipschitz functions

We start by introducing a useful class of functions between metric spaces whose members are always continuous.

Definition 3.1 *Lipschitz function*

Let (X, d) and (Y, e) be metric spaces.

A function $f\colon X \to Y$ is a **Lipschitz function** if there is a non-negative real number M such that, for each $a, b \in X$,

$$e(f(a), f(b)) \leq Md(a, b).$$

The real number M is known as a **Lipschitz constant for** f.

These functions are named after the German mathematician Rudolf Lipschitz (1832–1903).

Remarks

1. When we wish to emphasise the particular metrics d and e on X and Y, we say that f is a (d, e)-*Lipschitz function*.
2. If $(X, d) = (Y, e) = (\mathbb{R}, d^{(1)})$, then the definition reads:

   A function $f\colon \mathbb{R} \to \mathbb{R}$ is a Lipschitz function if there is a non-negative real number M such that, for each $a, b \in \mathbb{R}$,

   $$|f(b) - f(a)| \leq M|b - a|.$$

   By using the Mean Value Theorem it can be shown that $f(x) = \sin x$ satisfies

   $$|\sin b - \sin a| \leq |b - a|$$

   for each $a, b \in \mathbb{R}$; thus the sine function is an example of a $(d^{(1)}, d^{(1)})$-Lipschitz function (with Lipschitz constant $M = 1$).

   See the *Handbook*.
3. The solution of Worked Exercise 2.6 shows that the shift map on $\mathbf{C}$ is a $(d_{\mathbf{C}}, d_{\mathbf{C}})$-Lipschitz function.

The form of the inequality given in the definition is ready-made for proving that Lipschitz functions are continuous.

Theorem 3.2

Let (X, d) and (Y, e) be metric spaces.

If a function $f\colon X \to Y$ is (d, e)-Lipschitz, then it is (d, e)-continuous.

Proof Fix $a \in X$ and let (a_n) be a sequence in X that d-converges to a. We must show that $f(a_n) \to f(a)$ for the metric e as $n \to \infty$.

Since f is Lipschitz, there is a non-negative real number M such that for each $a, b \in X$,

$$e(f(a), f(b)) \leq Md(a, b).$$

In particular, for each $n \in \mathbb{N}$,

$$e(f(a_n), f(a)) \leq Md(a_n, a).$$

Since $a_n \xrightarrow{d} a$ as $n \to \infty$, we know that $(d(a_n, a))$ is a real null sequence, and then the Multiple Rule (for null sequences) gives $Md(a_n, a) \to 0$.

We now apply the Squeeze Rule to deduce that

$$e(f(a_n), f(a)) \to 0 \text{ as } n \to \infty;$$

that is, $(f(a_n))$ e-converges to $f(a)$.

We conclude that f is continuous at a. Since a is an arbitrary point of X, f is continuous on X. ■

We showed that f is continuous in Worked Exercise 2.3 of Chapter 13.

Thus all Lipschitz functions are continuous – but not all continuous functions are Lipschitz. For example, for the Euclidean metric on $\mathbb{R}$, the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is continuous, but not Lipschitz. For if f were Lipschitz, then there would be a constant $M \geq 0$ such that, for each $a, b \in \mathbb{R}$,

$$|b^2 - a^2| \leq M|b - a|,$$

$|b^2 - a^2| = |b + a|\,|b - a|$.

and so $|b + a| \leq M$ for $a \neq b$. In particular, if $b = 0$ then we would have $|a| \leq M$ for each $a \in \mathbb{R}$, which is clearly false. Hence f is not Lipschitz.

This means that being Lipschitz is more restrictive than being continuous but, despite this, many commonly occurring functions are Lipschitz. In particular, it is often the case that the easiest way to prove that a function is continuous is to prove that it is Lipschitz.

Worked Exercise 3.3

Fix $n \in \mathbb{N}$ and let $X = A^n$ where $A = \{0, 1\}$. Let $f\colon \mathbf{C} \to X$ be given by

$$f((a_1, a_2, a_3, \ldots)) = (a_1, a_2, a_3, \ldots, a_n).$$

(So f maps the infinite binary sequence $\mathbf{a}$ to its first n terms.) Show that f is $(d_{\mathbf{C}}, d_H)$-continuous on $\mathbf{C}$.

Solution

Fix $n \in \mathbb{N}$. We show that f is $(d_{\mathbf{C}}, d_H)$-Lipschitz and use Theorem 3.2.

Let $\mathbf{a} = (a_1, a_2, a_3, \ldots)$ and $\mathbf{b} = (b_1, b_2, b_3, \ldots)$ in $\mathbf{C}$.

We distinguish three cases:

1. If $\mathbf{a} = \mathbf{b}$, then $d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 0$ and $f(\mathbf{a}) = f(\mathbf{b})$ and so, in this case, we find
$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 0 = d_H(f(\mathbf{a}), f(\mathbf{b})).$$
2. If $\mathbf{a} \neq \mathbf{b}$, then suppose they first differ at the mth place. Then
$$d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 2^{-m}.$$
   (a) If $m > n$, then $f(\mathbf{a}) = f(\mathbf{b})$ and so $d_H(f(\mathbf{a}), f(\mathbf{b})) = 0$, giving
   $$0 = d_H(f(\mathbf{a}), f(\mathbf{b})) \leq d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}).$$
   (b) If, instead, $m \leq n$, then it is certainly the case that $d_H(f(\mathbf{a}), f(\mathbf{b})) \leq n$, since the maximum Hamming distance between words of length n is n. But $d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = 2^{-m} \geq 2^{-n}$. Hence if we let $M = n2^n$, a constant that does not depend on the choice of $\mathbf{a}$ and $\mathbf{b}$, then we have
   $$d_H(f(\mathbf{a}), f(\mathbf{b})) \leq n = n\left(2^n \cdot 2^{-n}\right) \leq n2^n \cdot 2^{-m} = Md_{\mathbf{C}}(\mathbf{a}, \mathbf{b}).$$

Thus in all cases we have

$$d_H(f(\mathbf{a}), f(\mathbf{b})) \leq M d_{\mathbf{C}}(\mathbf{a}, \mathbf{b})$$

and so f is $(d_{\mathbf{C}}, d_H)$-Lipschitz (with Lipschitz constant $M = n2^n$). We conclude by Theorem 3.2 that f is $(d_{\mathbf{C}}, d_H)$-continuous on $\mathbf{C}$.

Exercise 3.1

Fix $n \in \mathbb{N}$ and let $X = A^n$ where $A = \{0, 1\}$. Let $f\colon X \to \mathbf{C}$ be given by

$$f((a_1, a_2, \ldots, a_n)) = (a_1, a_2, a_3, \ldots, a_n, 0, 0, 0, \ldots).$$

Show that f is $(d_H, d_{\mathbf{C}})$-continuous on X.

3.2 Equivalence of metrics

We have defined several different notions of product metric for the plane, one of which is the usual definition of Euclidean distance on the plane. This leads to a natural question: if $f\colon \mathbb{R}^2 \to Y$ is a function from $\mathbb{R}^2$ to the metric space (Y, e), does its continuity depend upon which of the three metrics e_1, e_2, e_∞ we used for creating a metric on $\mathbb{R}^2$ (when applied to $d^{(1)}$)? It turns out that it does not: we obtain the same convergent sequences and so, as a result, the same continuous functions. The basic reason why this is the case is that all three metrics are comparable to each other, in that there are simple inequalities that relate any two of them.

We are assuming that each of e_1, e_2 and e_∞ is formed from taking the product of $(\mathbb{R}, d^{(1)})$ with itself – if we change the metric on one of the component spaces for one of the products, there is no reason to suppose that e_1, e_2 and e_∞ will be similar.

Worked Exercise 3.4

Show that for each $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$, if e_1, e_2 and e_∞ are formed by taking the product of $(\mathbb{R}, d^{(1)})$ with itself, then:

(a) $e_2(\mathbf{x}, \mathbf{y}) \leq e_1(\mathbf{x}, \mathbf{y})$

(b) $e_\infty(\mathbf{x}, \mathbf{y}) \leq e_2(\mathbf{x}, \mathbf{y})$

(c) $e_1(\mathbf{x}, \mathbf{y}) \leq 2e_\infty(\mathbf{x}, \mathbf{y})$.

Solution

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

(a) We use the fact that if a and b are non-negative real numbers, then $a^2 + b^2 \leq a^2 + 2ab + b^2 = (a + b)^2$, and so $\sqrt{a^2 + b^2} \leq a + b$.

$$\begin{aligned} e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{|y_1 - x_1|^2 + |y_2 - x_2|^2} \\ &\leq |y_1 - x_1| + |y_2 - x_2| = e_1(\mathbf{x}, \mathbf{y}). \end{aligned}$$

(b) $e_\infty(\mathbf{x},\mathbf{y}) = \max\{|y_1 - x_1|,\ |y_2 - x_2|\}$
$\leq \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} = e_2(\mathbf{x},\mathbf{y})$.

(c) $e_1(\mathbf{x},\mathbf{y}) = |y_1 - x_1| + |y_2 - x_2|$
$\leq \max\{|y_1 - x_1|, |y_2 - x_2|\} + \max\{|y_1 - x_1|, |y_2 - x_2|\}$
$= 2\max\{|y_1 - x_1|, |y_2 - x_2|\} = 2e_\infty(\mathbf{x},\mathbf{y})$.

These inequalities imply that for each $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$,

$$e_2(\mathbf{x},\mathbf{y}) \leq e_1(\mathbf{x},\mathbf{y}) \leq 2e_2(\mathbf{x},\mathbf{y}), \tag{1}$$
$$e_\infty(\mathbf{x},\mathbf{y}) \leq e_2(\mathbf{x},\mathbf{y}) \leq 2e_\infty(\mathbf{x},\mathbf{y}), \tag{2}$$
$$\tfrac{1}{2}e_1(\mathbf{x},\mathbf{y}) \leq e_\infty(\mathbf{x},\mathbf{y}) \leq e_1(\mathbf{x},\mathbf{y}). \tag{3}$$

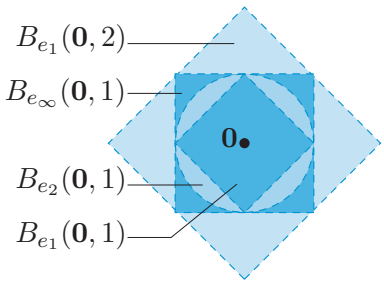


Figure 3.1 Unit open balls for e_1, e_2 and e_∞

One way to understand these inequalities is to consider the relationships between the unit open balls for each of the metrics e_1, e_2 and e_∞. Figure 3.1 shows that $B_{e_1}(\mathbf{0}, 1)$ fits inside $B_{e_2}(\mathbf{0}, 1)$, which fits inside $B_{e_\infty}(\mathbf{0}, 1)$. Moreover, $B_{e_\infty}(\mathbf{0}, 1)$ fits inside a scaled version of $B_{e_1}(\mathbf{0}, 1)$, namely the ball $B_{e_1}(\mathbf{0}, 2)$. By translating and rescaling, we see that versions of these inclusions hold for open balls with any given centre $\mathbf{a}$ and radius r.

Metrics that are related in this way are said to be *(metrically) equivalent*.

Definition 3.5 *Equivalent metrics*

Let d_1 and d_2 be two metrics defined on a set X. The metrics d_1 and d_2 are (metrically) **equivalent** if there are positive real numbers m and M such that for each $x, y \in X$,

$$md_1(x,y) \leq d_2(x,y) \leq Md_1(x,y). \tag{4}$$

Remarks

1. Notice that the inequalities making up equation (4) can be rearranged to give

   $$\tfrac{1}{M}d_2(x,y) \leq d_1(x,y) \leq \tfrac{1}{m}d_2(x,y)$$

   and so there is no asymmetry in the definition: saying that d_1 is (metrically) equivalent to d_2 is the same as saying that d_2 is (metrically) equivalent to d_1.
2. Inequalities (1)–(3) indicate that the metrics e_1, e_2 and e_∞ on $\mathbb{R}^2$ are all equivalent to each other.
3. If we fix the set X and consider the set of all possible metrics for X, then 'metrical equivalence' gives an equivalence relation on the set of metrics for X.

Exercise 3.2

Show that e_1 is equivalent on $\mathbb{R}^2$ to the metric e for $\mathbb{R}^2$ given by

$$e(\mathbf{x}, \mathbf{y}) = |y_1 - x_1| + 2|y_2 - x_2| \text{ for } \mathbf{x} = (x_1, x_2),\ \mathbf{y} = (y_1, y_2) \in \mathbb{R}^2.$$

You may assume that e is a metric on the plane.

The argument used to obtain inequalities (1)–(3) works not only for the product metrics on $\mathbb{R}^2$ but also for the metrics e_1, e_2 and e_∞ defined for a general product $X_1 \times X_2$. We omit the details, which are not very different from those for $\mathbb{R}^2$, and simply record the result.

Theorem 3.6

Let (X_1, d_1) and (X_2, d_2) be metric spaces, and let $X = X_1 \times X_2$. Then the three product metrics e_1, e_2 and e_∞ formed from d_1 and d_2 are equivalent.

We now explore why the notion of equivalence is useful: equivalent metrics have the same convergent sequences.

Proposition 3.7

Let d_1 and d_2 be equivalent metrics defined on the set X. Then a sequence (x_n) converges with limit x in (X, d_1) if, and only if, the same sequence (x_n) converges with the same limit x in (X, d_2).

Proof If the sequence (x_n) d_1-converges to x in X then $(d_1(x_n, x))$ is a real null sequence.

Since d_1 and d_2 are equivalent metrics, there is a fixed positive real number M such that for each $n \in \mathbb{N}$, $0 \leq d_2(x_n, x) \leq M d_1(x_n, x)$.

Thus $(d_2(x_n, x))$ is also a real null sequence by the Squeeze Rule and so (x_n) d_2-converges to x in X.

Conversely, if the sequence (x_n) d_2-converges to a limit x in X, then $(d_2(x_n, x))$ is a null sequence.

Since d_1 and d_2 are equivalent metrics, there is a fixed positive real number m such that for each $n \in \mathbb{N}$, $0 \leq d_1(x_n, x) \leq \frac{1}{m} d_2(x, y)$.

Thus $(d_1(x_n, x))$ is a real null sequence by the Squeeze Rule and so (x_n) d_1-converges with limit x in X. ■

This result has an important corollary.

Theorem 3.8

Let (Y, e) be a metric space and let d_1 and d_2 be metrically equivalent metrics on a set X.

Then

(a) a function $f\colon X \to Y$ is (d_1, e)-continuous at a point $x \in X$ if, and only if, f is (d_2, e)-continuous at x; and

(b) a function $g\colon Y \to X$ is (e, d_1)-continuous at a point $y \in Y$ if, and only if, g is (e, d_2)-continuous at y.

Exercise 3.3

(a) Prove the first part of Theorem 3.8.

(b) Prove the second part of Theorem 3.8.

Since Theorem 3.6 tells us that the metrics e_1, e_2 and e_∞ are equivalent on any product $X_1 \times X_2$, the following result follows immediately.

Theorem 3.9

Let (X_1, d_1), (X_2, d_2) and (Y, e) be metric spaces, let $X = X_1 \times X_2$ and let e_1, e_2 and e_∞ be the three different product metrics on X (determined by d_1 and d_2).

If $f\colon X \to Y$ then, for $j, k \in \{1, 2, \infty\}$, f is (e_j, e)-continuous at $x \in X$ if, and only if, f is (e_k, e)-continuous at x.

If $g\colon Y \to X$ then, for $j, k \in \{1, 2, \infty\}$, g is (e, e_j)-continuous at $y \in Y$ if, and only if, g is (e, e_k)-continuous at y.

Remarks

1. Suppose we have a function f from $X_1 \times X_2$ with one of the product metrics e_1, e_2 or e_∞. Theorem 3.9 tells us that we can test f for continuity by using any one of e_1, e_2 or e_∞, and so in practice we choose the one that is simplest to calculate with.
2. There is a natural extension of this result to the case when X is an n-fold product of metric spaces $X_1, X_2 \ldots, X_n$; we omit the details.

Projection functions and continuity

In Subsection 1.4 of Chapter 14, we saw that we can show a function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is continuous by showing that both $p_1 \circ f\colon \mathbb{R}^2 \to \mathbb{R}$ and $p_2 \circ f\colon \mathbb{R}^2 \to \mathbb{R}$ are continuous, where $p_1\colon \mathbb{R}^2 \to \mathbb{R}$ and $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ are the projection functions defined by $p_1(x_1, x_2) = x_1$ and $p_2(x_1, x_2) = x_2$. We can extend the definitions of projection functions to a general product space, and a similar result about continuity holds.

Definition 3.10 *Projection functions on a space*

Let X_1 and X_2 be sets and let $X = X_1 \times X_2$. The **projection functions** are the functions $p_1\colon X \to X_1$ and $p_2\colon X \to X_2$ given by

$$p_1(x_1, x_2) = x_1 \text{ and } p_2(x_1, x_2) = x_2.$$

As with the projection functions in Chapter 13, these functions are continuous.

Theorem 3.11

Let (X_1, d_1) and (X_2, d_2) be metric spaces and let e be one of the product metrics e_1, e_2 or e_∞ for $X_1 \times X_2$.

Then p_1 is (e, d_1)-continuous and p_2 is (e, d_2)-continuous, where p_1 and p_2 are the projection functions from X to X_1 and X_2 respectively.

Proof Once we have proved the result for any one of e_1, e_2 or e_∞, the others follow immediately from Theorem 3.9.

We prove the result for e_∞ as this turns out to be the simplest to work with.

We show that p_1 is (e_∞, d_1)-continuous; the proof that p_2 is (e_∞, d_2)-continuous is similar.

Fix $\mathbf{a} = (a_1, a_2), \mathbf{b} = (b_1, b_2) \in X$. We find an upper bound on $d_1(p_1(\mathbf{a}), p_1(\mathbf{b}))$ in terms of $e_\infty(\mathbf{a}, \mathbf{b})$:

$$\begin{aligned} d_1(p_1(\mathbf{a}), p_1(\mathbf{b})) &= d_1(a_1, b_1) \\ &\leq \max\{d_1(a_1, b_1), d_2(a_2, b_2)\} = e_\infty(\mathbf{a}, \mathbf{b}). \end{aligned}$$

Thus p_1 is an (e_∞, d_1)-Lipschitz function and so, by Theorem 3.2, p_1 is (e_∞, d_1)-continuous on X. ■

We can now derive the promised continuity result. Since its proof is similar to that of Proposition 1.19 of Chapter 14, we omit it.

Theorem 3.12

Let (Y, d), (X_1, d_1) and (X_2, d_2) be metric spaces, and let (X, e) be the metric space where $X = X_1 \times X_2$ and e is one of the three product metrics e_1, e_2 or e_∞ formed from d_1 and d_2.

Then a function $f: Y \to X_1 \times X_2$ is (d, e)-continuous at $a \in Y$ if, and only if, $p_1 \circ f: Y \to X_1$ is (d, d_1)-continuous at a and $p_2 \circ f: Y \to X_2$ is (d, d_2)-continuous at a.

Theorems 3.11 and 3.12 extend in a natural way to n-fold products of metric spaces $X_1 \times \cdots \times X_n$ once one defines appropriate projection functions $p_i: X_1 \times \cdots \times X_n \to X_i$ by $p_i(x_1, x_2, \ldots, x_n) = x_i$. We omit the details.

Worked Exercise 3.13

Consider the metric spaces $(\mathbf{C}, d_\mathbf{C})$ and $(\mathbb{R}, d^{(1)})$. Define the function $F: \mathbf{C} \to \mathbb{R} \times \mathbf{C}$ by

$$F(\mathbf{a}) = \left(\sum_{n=1}^{\infty} 2^{-n} a_n, \mathbf{a} \right).$$

Notice that the sum in the first component of the definition of $F(\mathbf{a})$ always exists and is finite since each term a_n is either zero or one.

Show that F is $(d_\mathbf{C}, e)$-continuous, where e denotes any of e_1, e_2 or e_∞.

Solution

We consider the composites of the projection functions with F, that is, $p_1 \circ F: \mathbf{C} \to \mathbb{R}$ and $p_2 \circ F: \mathbf{C} \to \mathbf{C}$.

Now, for $\mathbf{a} = (a_1, a_2, a_3, \ldots) \in \mathbf{C}$,

$$(p_1 \circ F)(\mathbf{a}) = \sum_{n=1}^{\infty} 2^{-n} a_n.$$

We show that $p_1 \circ F$ is $(d_\mathbf{C}, d^{(1)})$-Lipschitz. So suppose that $\mathbf{a} = (a_1, a_2, a_3, \ldots), \mathbf{b} = (b_1, b_2, b_3, \ldots) \in \mathbf{C}$. If they are identical, then

$$d^{(1)}((p_1 \circ F)(\mathbf{a}), (p_1 \circ F)(\mathbf{b})) = 0 = d_\mathbf{C}(\mathbf{a}, \mathbf{b}).$$

So suppose instead that they first differ at the mth place. Then $d_\mathbf{C}(\mathbf{a}, \mathbf{b}) = 2^{-m}$ and

$$
\begin{aligned}
&d^{(1)}((p_1 \circ F)(\mathbf{a}), (p_1 \circ F)(\mathbf{b})) \\
&\qquad = \left| \sum_{n=1}^{\infty} 2^{-n}(a_n - b_n) \right| \\
&\qquad \leq \left| \sum_{n=m}^{\infty} 2^{-n}(a_n - b_n) \right|, \text{ since } a_n = b_n \text{ for } n < m \\
&\qquad \leq 2^{1-m} \\
&\qquad = 2d_{\mathbf{C}}(\mathbf{a}, \mathbf{b}).
\end{aligned}
$$

Hence $p_1 \circ F$ is $(d_{\mathbf{C}}, d^{(1)})$-Lipschitz (with Lipschitz constant 2), and so by Theorem 3.2 it is $(d_{\mathbf{C}}, d^{(1)})$-continuous.

Also,

$$(p_2 \circ F)(\mathbf{a}) = \mathbf{a},$$

and so

$$d_{\mathbf{C}}((p_2 \circ F)(\mathbf{a}), (p_2 \circ F)(\mathbf{b})) = d_{\mathbf{C}}(\mathbf{a}, \mathbf{b})$$

for each $\mathbf{a}, \mathbf{b} \in \mathbf{C}$. Thus $p_2 \circ F$ is a $(d_{\mathbf{C}}, d_{\mathbf{C}})$-Lipschitz function (with Lipschitz constant 1), and so by Theorem 3.2 it is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous.

Since both $p_1 \circ F$ and $p_2 \circ F$ are continuous, Theorem 3.12 implies that F is $(d_{\mathbf{C}}, e)$-continuous, where e is any of e_1, e_2 or e_∞.

Exercise 3.4

Consider the metric spaces $(\mathbf{C}, d_{\mathbf{C}})$ and $(\mathbb{R}, d^{(1)})$. Define the function $f\colon \mathbf{C} \to \mathbb{R} \times \mathbf{C}$ by

$$f(\mathbf{a}) = (a_1, \sigma(\mathbf{a})).$$

(a) Evaluate $f(\mathbf{a})$ when:

(i) $\mathbf{a} = (0, 0, 0, 0, 0, 0, \ldots)$

(ii) $\mathbf{a} = (0, 1, 0, 1, 0, 1, \ldots)$.

(b) Show that f is $(d_{\mathbf{C}}, e)$-continuous, where e denotes any of e_1, e_2 or e_∞.

Hint: you may assume that $g\colon \mathbf{C} \to \mathbb{R}$ given by $g(\mathbf{a}) = a_1$ is $(d_{\mathbf{C}}, d^{(1)})$-continuous.

3.3 Convergence in product spaces

We finish this section by taking a brief look at the relationship between convergence of sequences in product spaces and convergence of the corresponding sequences in the component spaces that make up the product.

Let (X, d_1) and (Y, d_2) be metric spaces and let e be one of the product metrics e_1, e_2 or e_∞ on $X \times Y$.

A sequence in $X \times Y$ can be written as a sequence of ordered pairs, $((x_n, y_n))$ – that is, for each $n \in \mathbb{N}$, the ordered pair (x_n, y_n) belongs to $X \times Y$, with $x_n \in X$ and $y_n \in Y$. For example, if $X = Y = \mathbb{R}$, then

$$(1, -1),\ (2, 0),\ (3, 1),\ \ldots,\ (n, n-2),\ \ldots$$

is a sequence in $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.

Definition 3.14 *Component sequences*

Let X and Y be sets and let $((x_n, y_n))$ be a sequence in $X \times Y$. The sequences (x_n) in X and (y_n) in Y are the **component sequences** or **coordinate sequences** of the **product sequence** $((x_n, y_n))$.

The component sequences for the above example are $(1, 2, 3, \ldots, n, \ldots)$ and $(-1, 0, 1, \ldots, n-2, \ldots)$.

Note that the component sequences (x_n) and (y_n) are the images of the original sequence $((x_n, y_n))$ under the projection maps $p_1 \colon X \times Y \to X$ and $p_2 \colon X \times Y \to Y$ given by $p_1((x, y)) = x$ and $p_2((x, y)) = y$.

Since Theorem 3.11 tells us that the projection map p_i is (e, d_i)-continuous for $i = 1, 2$, we can immediately conclude that if $((x_n, y_n))$ e-converges to $(x, y) \in X \times Y$, then $x_n \xrightarrow{d_1} x$ and $y_n \xrightarrow{d_2} y$ as $n \to \infty$. The following result shows that the converse is also true.

Theorem 3.15

Let (X, d_1) and (Y, d_2) be metric spaces and let e be one of the three product metrics e_1, e_2 and e_∞ for $X \times Y$.

Suppose that $(x, y) \in X \times Y$ and $((x_n, y_n))$ is a sequence in $X \times Y$. Then

$$(x_n, y_n) \xrightarrow{e} (x, y) \quad \text{if, and only if,} \quad x_n \xrightarrow{d_1} x \text{ and } y_n \xrightarrow{d_2} y \text{ as } n \to \infty.$$

Proof We have already noted that if $(x_n, y_n) \xrightarrow{e} (x, y)$, then both $x_n \xrightarrow{d_1} x$ and $y_n \xrightarrow{d_2} y$.

It remains to show the converse. So suppose that $x_n \xrightarrow{d_1} x$ and $y_n \xrightarrow{d_2} y$ as $n \to \infty$. Then both $(d_1(x_n, x))$ and $(d_2(y_n, y))$ are (real) null sequences. But for the product metric e_1,

$$e_1((x_n, y_n), (x, y)) = d_1(x_n, x) + d_2(y_n, y)$$

and so, by the Sum Rule, $(e_1((x_n, y_n), (x, y)))$ is a null sequence, meaning that $((x_n, y_n))$ converges in $(X \times Y, e_1)$ to (x, y).

We now use the equivalence of the three metrics e_1, e_2 and e_∞ and Proposition 3.7 to conclude that $((x_n, y_n))$ converges to (x, y) in $(X \times Y, e)$, where e is any of the three product metrics. ■

By induction, we can prove an analogous result for the product of k metric spaces.

Let (X_1, d_1), (X_2, d_2), …, (X_k, d_k) be metric spaces, and let e be a product metric on the set $X_1 \times X_2 \times \cdots \times X_k$. A sequence $((x_{1,n}, x_{2,n}, \ldots, x_{k,n}))_{n \in \mathbb{N}}$ in the product set $X_1 \times X_2 \times \cdots \times X_k$ converges to $(x_1, x_2, \ldots, x_k)$ with respect to the product metric e if, and only if, each component sequence converges for its own metric: $x_{j,n} \xrightarrow{d_j} x_j$ as $n \to \infty$, for $j = 1, 2, \ldots, k$.

Exercise 3.5

Consider the sequence $((\frac{1}{n}, \frac{1}{n^2}))$ in $\mathbb{R}^2$, and let e be the product metric obtained by taking the product of the Euclidean metric with the discrete metric. Does the sequence converge for e and, if so, to what?

$e((x_1, x_2), (y_1, y_2)) = d^{(1)}(x_1, y_1) + d_0(x_2, y_2)$.

4 Spaces of functions

In this section we continue our study of (real-valued) continuous functions by looking at two different ways of defining convergence for sequences of functions – *pointwise convergence* and *uniform convergence*.

Once we have these notions we can ask whether, if the functions $f_n \colon X \to Y$ are continuous (with respect to some metrics on X and Y), the limit function, if it exists, is necessarily continuous. We show that the notion of pointwise convergence does not guarantee this, but that uniform convergence does.

4.1 Pointwise convergence

We first discuss the simplest way of defining convergence for sequences of real-valued functions – pointwise convergence.

There is no assumption here that A has a metric defined on it.

If A is some fixed set and $f_n\colon A \to \mathbb{R}$ for each $n \in \mathbb{N}$ then, for each $x \in A$, $(f_n(x))$ is a sequence of real numbers. Pointwise convergence involves looking separately at each x in the set A and investigating whether the corresponding sequence of real numbers $(f_n(x))$ converges.

Definition 4.1 *Pointwise convergence*

Let A be a set.

A sequence (f_n) of functions $f_n\colon A \to \mathbb{R}$ **converges pointwise** to the function $f\colon A \to \mathbb{R}$ if, for each $x \in A$, the sequence of real numbers $(f_n(x))$ converges to $f(x)$.

That is, $(|f_n(x) - f(x)|)_{n\in\mathbb{N}}$ is a real null sequence.

If a sequence (f_n) of functions converges pointwise to a function f, then f is called the **pointwise limit** of the sequence (f_n).

We write $f_n \to f$ pointwise as $n \to \infty$, $f_n \to_p f$ as $n \to \infty$ or simply $f_n \to f$ pointwise.

Remarks

1. We have not assumed anything about the continuity of the individual functions f_n. Indeed, since A need not have a metric defined for it, it may not even make sense to talk about the continuity of the functions f_n.
2. If it exists, then we know that the pointwise limit is unique, because for each $x \in A$, the sequence $(f_n(x))$ is a sequence of real numbers, and the convergence of a sequence of real numbers corresponds to convergence in $\mathbb{R}$ with respect to the Euclidean metric. In particular, limits (when they exist) are unique.

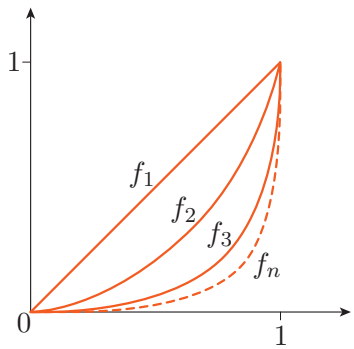


Figure 4.1 The functions f_n

Worked Exercise 4.2

For each $n \in \mathbb{N}$, let $f_n\colon [0,1] \to \mathbb{R}$ be defined by $f_n(x) = x^n$. Find the pointwise limit of the sequence (f_n).

Solution

Since $x^n \to 0$ for each $x \in [0,1)$ but $x^n \to 1$ when $x = 1$, we need to consider the point $x = 1$ separately.

For a fixed $x \in [0,1)$, the sequence of values of the functions at x is $(f_n(x)) = (x^n)$, and this real sequence has limit 0 as $n \to \infty$ (Figure 4.1). Thus, for $x \in [0,1)$, $f_n(x) \to 0$ as $n \to \infty$.

For $x = 1$ and all $n \in \mathbb{N}$, $f_n(1) = 1$, so $f_n(1) \to 1$ as $n \to \infty$.

Hence the pointwise limit of the sequence (f_n) is the function $f\colon [0,1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x = 1. \end{cases}$$

Notice that each of the functions f_n in Worked Exercise 4.2 is continuous (with respect to the Euclidean metric) – see Figure 4.1 – and yet the limit function is not (see Figure 4.2). This illustrates the fundamental disadvantage of pointwise convergence – it does not generally preserve any 'nice' properties of the functions f_n in the limit. We will see further examples of this shortly, but for now the result is highlighted below.

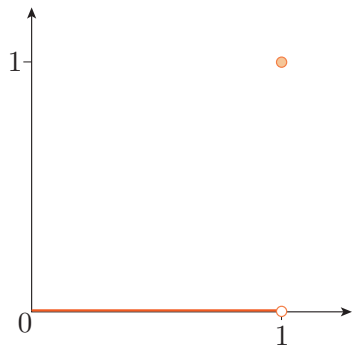


Figure 4.2 The limit function

The pointwise limit of a sequence of continuous functions need not be continuous.

Exercise 4.1

For each $n \in \mathbb{N}$, let $f_n\colon [0,\infty) \to \mathbb{R}$ be defined by

$$f_n(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq \frac{1}{2^n}, \\ -\frac{1}{2^n}, & \text{if } \frac{1}{2^n} < x < \infty. \end{cases}$$

Find the pointwise limit of the sequence of functions (f_n).

Exercise 4.2

For each $n \in \mathbb{N}$, let $f_n\colon [0,1] \to \mathbb{R}$ be defined by

$$f_n(x) = \begin{cases} 4n^2x, & \text{if } 0 \leq x \leq \frac{1}{2n}, \\ -4n^2x + 4n, & \text{if } \frac{1}{2n} < x < \frac{1}{n}, \\ 0, & \text{if } \frac{1}{n} \leq x \leq 1. \end{cases}$$

(a) Sketch the graphs of f_1, f_2 and f_3.

(b) Find the pointwise limit f of the sequence (f_n), and calculate $\int_0^1 f(x)\,dx$.

(c) Find the limit of the real sequence $(\int_0^1 f_n(x)\,dx)_{n\in\mathbb{N}}$.

(d) What conclusion, if any, can be drawn about the relation between integration and pointwise convergence?

4.2 Uniform convergence

We have seen that pointwise convergence of functions need not preserve continuity: a sequence of continuous functions may have a discontinuous pointwise limit. In this subsection, we introduce a type of convergence that does preserve continuity – *uniform convergence*.

You can learn more about Augustin-Louis Cauchy (1789–1857) in the *History Reader*.

Historically, the definition of uniform convergence arose out of a mistake made by Cauchy in his *Lectures on Analysis*, published in 1821. Cauchy was considering an infinite *series* of functions $u_i\colon [0,1] \to \mathbb{R}$, and made the following assertion:

> if the functions u_i are all continuous and if the sequence of partial sums $s_n\colon [0,1] \to \mathbb{R}$ given by
>
> $$s_n(x) = \sum_{i=1}^{n} u_i(x)$$
>
> converges pointwise to a limit function s, then s is also continuous.

The function s_n is continuous, since it is a *finite* sum of continuous functions.

Here $s = \sum_{i=1}^{\infty} u_i$.

In the last subsection, we saw that pointwise convergence does not preserve continuity in general, and it is not difficult to find functions u_n for which the pointwise limit of the partial sums exists and is *not* continuous, meaning that Cauchy's assertion is incorrect. The error was first noticed in 1826 by Abel, who presented examples of sequences of (continuous) partial sums whose pointwise limit is discontinuous.

The *History Reader* says more about Niels Henrik Abel (1802–1829) and his investigation of the convergence of sequences and series.

Let us see where Cauchy's mistake lies. He wrote

$$s(x) = s_n(x) + \sum_{i=n+1}^{\infty} u_i(x),$$

and reasoned as follows. Since s_n converges to s as $n \to \infty$, given any $\varepsilon > 0$ we can choose an $N \in \mathbb{N}$, depending on ε, so that for each $n > N$, the 'tail' of the series is less than ε:

$$\left| \sum_{i=n+1}^{\infty} u_i(x) \right| < \varepsilon.$$

Hence, he wrote, $s(x)$ is a finite sum of continuous functions plus a correction that is less than ε, and so s is continuous.

The issue is that, in general, N depends upon both ε *and* x, and so the whole chain of reasoning just given breaks down. But if the functions u_i are such that we *can* choose an N independently of x, then the result follows. In this case, the same N works *simultaneously* for all x.

We showed in Worked Exercise 4.2 that the limit function $f\colon [0,1] \to \mathbb{R}$ is given by

$$f(x) = \begin{cases} 0, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x = 1. \end{cases}$$

To see what can go wrong, let us reconsider the convergence of the sequence (f_n) where $f_n\colon [0,1] \to \mathbb{R}$ is given by $f_n(x) = x^n$. We know that the limit function f is discontinuous with respect to the Euclidean metric, and so whatever goes wrong with pointwise convergence should be discernible from this sequence.

Suppose that $0 < \varepsilon < 1$ and $x \in (0,1)$. For what value of n does x^n become less than ε? Solving $x^r = \varepsilon$ for r, we find that $r = \frac{\log \varepsilon}{\log x}$. As anticipated, $r = r(\varepsilon, x)$ depends on both x and ε. Thus, if $N(\varepsilon, x) \in \mathbb{N}$ is chosen so that $N(\varepsilon, x) \geq r(\varepsilon, x)$, then $x^n < \varepsilon$ for each $n > N(\varepsilon, x)$.

The crucial question is whether there is an integer $N(\varepsilon)$, depending on ε but not x, that can serve for a 'uniform rate' of decrease of the x^n over the entire domain $[0,1]$.

$N(\varepsilon)$ needs to be at least as large as all of the numbers $N(\varepsilon, x)$.

The answer is *no*, for as x approaches 1 from below, $N(\varepsilon, x)$ grows without limit; that is,

$$\sup\{N(\varepsilon, x) : x \in [0,1]\} \geq \sup\{\tfrac{\log \varepsilon}{\log x} : x \in (0,1)\} = \infty.$$

The hope is that if we require this *not* to happen – that is, if the set $\{N(\varepsilon, x) : x \in [0,1]\}$ is bounded above for each $\varepsilon > 0$ – then we obtain a 'uniform control', and the limit function of a sequence of continuous functions *is* continuous. This is the idea underlying the notion of *uniform convergence*.

Before we define uniform convergence, we need to extend our definition of a bounded function to situations when the domain may not be a subset of $\mathbb{R}$.

Definition 4.3 *Bounded function*

Let A be a set. A function $f\colon A \to \mathbb{R}$ is **bounded** on A if there is an $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in A$.

This extends the definition of boundedness of a function, which you met in Definition 2.14 of Chapter 13.

We can now give the definition of uniform convergence.

Definition 4.4 *Uniform convergence*

Let A be a set. A sequence (f_n) of functions $f_n\colon A \to \mathbb{R}$ **converges uniformly** on A to the function $f\colon A \to \mathbb{R}$ if

(a) the function $f_n - f$ is bounded for each $n \in \mathbb{N}$, and

(b) the sequence (M_n) defined by

$$M_n = \sup\{|f_n(x) - f(x)| : x \in A\}$$

is a (real) null sequence.

Such a function f is called the **uniform limit** of the sequence (f_n).

We write either $f_n \to f$ uniformly as $n \to \infty$, $f_n \rightrightarrows f$ as $n \to \infty$ or simply $f_n \to f$ uniformly.

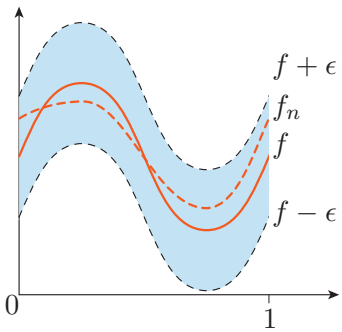


Figure 4.3 Uniform convergence

Remarks

1. The requirement that $f_n - f$ is bounded is imposed to ensure that M_n is a well-defined real number for each $n \in \mathbb{N}$.
2. As with the definition of pointwise convergence, we make no assumption as to whether there is a metric for the domain of the functions.
3. We must always specify on which domain the sequence converges uniformly. A sequence might converge uniformly on A but not on some larger set $B \supset A$. This is illustrated in Worked Exercise 4.6 and Exercise 4.4 (below).
4. There is a useful visual aid to understanding the meaning of uniform convergence when $A = [0, 1]$. Given $\varepsilon > 0$, we draw a 'sleeve' of total width 2ε symmetrically around the graph of f (Figure 4.3). If, for each $\varepsilon > 0$, there is an $N(\varepsilon) \in \mathbb{N}$ such that the graph of each function f_n for $n > N(\varepsilon)$ fits inside this sleeve over the whole of A, then the convergence is uniform. If there is at least one ε for which we cannot find such a sleeve, then the convergence is not uniform.

Before we look at some examples of uniform convergence, we note a useful lemma that relates uniform and pointwise convergence of sequences of functions: it states that uniform convergence implies pointwise convergence.

Lemma 4.5

Let A be a set and let (f_n) be a sequence of functions $f_n \colon A \to \mathbb{R}$.

If the sequence (f_n) converges uniformly on A to the function $f \colon A \to \mathbb{R}$, then (f_n) also converges pointwise to f.

Proof Suppose that (f_n) converges uniformly to f. We must show that for each $x \in A$, the real sequence $(f_n(x))$ converges to $f(x)$.

For each $n \in \mathbb{N}$, let

$$M_n = \sup\{|f_n(x) - f(x)| : x \in A\}$$

and note that, since f_n converges uniformly to f, $M_n < \infty$ for each n and (M_n) is a real null sequence.

But for each $n \in \mathbb{N}$ and each $x \in A$,

$$0 \leq |f_n(x) - f(x)| \leq M_n.$$

See Lemma 1.15 of Chapter 13.

Hence, by the Squeeze Rule for sequences, $(f_n(x) - f(x))_{n \in \mathbb{N}}$ is a null sequence, and so $f_n(x) \to f(x)$ as $n \to \infty$. That is, f_n converges to f pointwise. ■

This lemma tells us that if a sequence of functions converges uniformly to a function, then it also converges pointwise to the same function. Thus, in order to identify a possible uniform limit of a sequence of functions, it is enough to find the (unique) pointwise limit of the sequence. Having found this candidate for the uniform limit, we then have to verify whether or not it *is* the uniform limit. The following worked exercise illustrates the method.

Worked Exercise 4.6

For $n \in \mathbb{N}$, let $f_n : [0,1] \to \mathbb{R}$ be given by

$$f_n(x) = \tfrac{x}{n}.$$

Show that the sequence (f_n) converges uniformly on $[0,1]$, and find its uniform limit.

Solution

We begin by identifying the pointwise limit of the sequence. For each $x \in [0,1]$, $\frac{x}{n} \to 0$ as $n \to \infty$, and so the pointwise limit of the sequence (f_n) is the zero function $f : [0,1] \to \mathbb{R}$ given by $f(x) = 0$ for each $x \in [0,1]$.

Thus, by Lemma 4.5, our only candidate for the uniform limit of the sequence (f_n) is the function f. To show that f is the uniform limit of (f_n), we must show that the sequence (M_n), where M_n is defined by

$$M_n = \sup\{|f_n(x) - f(x)| : x \in [0,1]\},$$

is a null sequence.

To prove this, we observe that for each $x \in [0,1]$ and each $n \in \mathbb{N}$,

$$0 \leq |f_n(x) - f(x)| = |\tfrac{x}{n} - 0| = |\tfrac{x}{n}| \leq \tfrac{1}{n}.$$

Hence for each $n \in \mathbb{N}$,

$$0 \leq M_n = \sup\{|\tfrac{x}{n} - 0| : x \in [0,1]\} = \tfrac{1}{n}.$$

Since $(\frac{1}{n})$ is a basic null sequence, it follows that (M_n) is a null sequence. Hence $f_n \to f$ uniformly as $n \to \infty$.

Exercise 4.3

For $n \in \mathbb{N}$, let $f_n : [0,1] \to \mathbb{R}$ be given by

$$f_n(x) = 1 - (\tfrac{x}{n})^2.$$

Show that the sequence of functions (f_n) converges uniformly on $[0,1]$, and find its uniform limit.

Exercise 4.4

Let $f_n : \mathbb{R} \to \mathbb{R}$ be given by

$$f_n(x) = \frac{x}{n}.$$

Show that the sequence of functions (f_n) does not converge uniformly on $\mathbb{R}$.

We have seen in Worked Exercise 4.6 and Exercise 4.4 that the functions given by $f_n(x) = \frac{x}{n}$ converge uniformly on $[0, 1]$, but do *not* converge uniformly on the whole of $\mathbb{R}$. Exercise 4.4 also illustrates another point.

Even when a sequence converges pointwise to a *continuous* function, the convergence need not be uniform.

Uniform convergence and continuity

We now show that uniform convergence preserves continuity – that is, *the uniform limit of a sequence of continuous functions is also continuous.*

Recall that in order to consider continuity, we need a metric on the domain as well as a metric on the codomain $\mathbb{R}$. We choose to use the usual Euclidean metric for the codomain, $d^{(1)}(x, y) = |x - y|$.

Theorem 4.7

Let (X, d) be a metric space and fix $a \in X$.

For each $n \in \mathbb{N}$, suppose that $f_n : X \to \mathbb{R}$ is $(d, d^{(1)})$-continuous at a.

If the sequence of functions (f_n) converges uniformly on X to a function $f : X \to \mathbb{R}$, then f is $(d, d^{(1)})$-continuous at a.

Proof The proof of this result takes some care.

Let (X, d) be a metric space and fix $a \in X$. For each $n \in \mathbb{N}$, suppose that $f_n : X \to \mathbb{R}$ is $(d, d^{(1)})$-continuous at a and that the sequence of functions (f_n) converges uniformly to the function $f : X \to \mathbb{R}$.

Our goal is to show that f is $(d, d^{(1)})$-continuous at a; that is, if (x_m) is a sequence in X that d-converges to a, then $f(x_m) \to f(a)$ as $m \to \infty$.

Notice that the sequence is indexed by a different parameter from that used for the sequence of functions.

To do this, it is enough to show that for each $\varepsilon > 0$, there is $N \in \mathbb{N}$ for which $|f(x_m) - f(a)| < \varepsilon$ whenever $m > N$.

Fix $\varepsilon > 0$.

Since f_n converges uniformly to f as $n \to \infty$, we know that the sequence (M_n) defined by

$$M_n = \sup\{|f_n(x) - f(x)| : x \in X\}$$

is a real null sequence. Since this is the only information we have about the relationship between the functions f_n and f, it should not be a surprise that we use it in the proof.

Let us start by observing that for any term x_m of the sequence, we can use the Triangle Inequality for the modulus function to estimate that for each $m \in \mathbb{N}$ and $n \in \mathbb{N}$,

$$\begin{aligned} |f(x_m)-f(a)| &= |f(x_m) - f_n(x_m) + f_n(x_m) - f_n(a) + f_n(a) - f(a)| \\ &= |(f(x_m) - f_n(x_m)) + (f_n(x_m) - f_n(a)) + (f_n(a) - f(a))| \\ &\leq |f(x_m) - f_n(x_m)| + |f_n(x_m) - f_n(a)| + |f_n(a) - f(a)| \\ &\leq M_n + |f_n(x_m) - f_n(a)| + M_n \\ &= 2M_n + |f_n(x_m) - f_n(a)|. \end{aligned}$$

This splitting of an expression by adding and subtracting identical terms is a very useful technique in analysis.

However, we know that (M_n) is a real null sequence (of non-negative terms) and so, in particular, we can find $n \in \mathbb{N}$ for which $0 \leq M_n < \frac{\varepsilon}{3}$. But we also know that for this choice of n the function f_n is $(d, d^{(1)})$-continuous at a and so, since $x_m \xrightarrow{d} a$, there is $N \in \mathbb{N}$ for which $d^{(1)}(f_n(x_m), f_n(a)) < \frac{\varepsilon}{3}$ whenever $m > N$. We conclude that for this choice of N, whenever $m > N$,

$$|f(x_m) - f(a)| < 2 \times \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon,$$

as required.

Hence f is $(d, d^{(1)})$-continuous at a. ■

Theorems such as this one are very useful: they show that under certain circumstances, the limit of a sequence can have some of the properties of the terms of the sequence.

Exercise 4.5

For each $n \in \mathbb{N}$, let $f_n \colon [0, \infty) \to \mathbb{R}$ be given by

$$f_n(x) = \begin{cases} 1, & \text{if } x = 0, \\ \dfrac{\sin(nx)}{nx}, & \text{if } x > 0. \end{cases}$$

(a) Does this sequence converge pointwise?

(b) Does it converge uniformly? If it does converge, find its limit function.

Hint: $\lim_{y \to 0} \frac{\sin y}{y} = 1$.

5 The max metric on $C[0, 1]$

In this final section of the chapter, we introduce a metric for a particular set of functions that is based on ideas related to uniform convergence. In this example, the functions in question are the continuous functions from the closed interval $[0, 1]$ to $\mathbb{R}$, and the metric we define is known as the *max metric*.

Notation

We let $C[0, 1]$ denote the set consisting of all functions $f\colon [0, 1] \to \mathbb{R}$ that are continuous on $[0, 1]$ with respect to the Euclidean metrics on $[0, 1]$ and $\mathbb{R}$.

Thus for each $a \in [0, 1]$ and each sequence (x_k) in $[0, 1]$, if $|x_k - a| \to 0$ as $k \to \infty$, then $|f(x_k) - f(a)| \to 0$ as $k \to \infty$.

Remarks

1. The points of $C[0, 1]$ are continuous functions on $[0, 1]$ such as $f\colon [0, 1] \to \mathbb{R}$ given by $f(x) = 3x$ and $g\colon [0, 1] \to \mathbb{R}$ given by $g(x) = \sin x$, for $x \in [0, 1]$.
2. We define the set $C[a, b]$ of all continuous functions from the closed interval $[a, b]$ to $\mathbb{R}$ in a similar way.

These properties can be deduced from the combination rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$, together with the Restriction Rule (Proposition 2.8) from Chapter 13.

In discussing the max metric, we are going to need the following properties of functions in $C[0, 1]$, all of which can be easily deduced from corresponding results given in Chapter 13.

Properties of functions in $C[0, 1]$

Let f and g be real-valued continuous functions with domain $[0, 1]$, i.e. $f, g \in C[0, 1]$, and let $\lambda \in \mathbb{R}$. Then the following real-valued functions are all continuous on $[0, 1]$, i.e. they are all members of $C[0, 1]$:

Sum Rule	$f + g\colon [0, 1] \to \mathbb{R}$, defined by $(f + g)(x) = f(x) + g(x)$
Multiple Rule	$\lambda f\colon [0, 1] \to \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$
Product Rule	$fg\colon [0, 1] \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$
Modulus Rule	$\lvert f\rvert\colon [0, 1] \to \mathbb{R}$, defined by $\lvert f\rvert(x) = \lvert f(x)\rvert$.

A *real-valued function* is one with codomain $\mathbb{R}$.

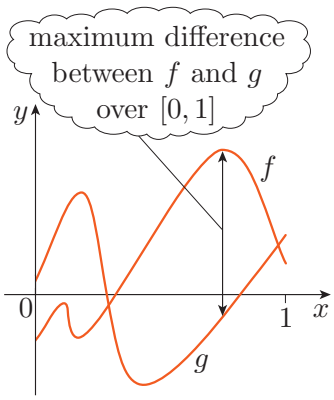


Figure 5.1 The max metric measures the maximum difference between two functions

The metric we define on $C[0, 1]$ is based on the idea of measuring the distance between two functions f and g in $C[0, 1]$ in terms of their maximum difference over the interval $[0, 1]$. This is given by

$$\max\{|g(x) - f(x)| : x \in [0, 1]\}$$

and is illustrated in Figure 5.1. It makes sense to talk of this quantity, because the map $x \mapsto |g(x) - f(x)|$ defines a continuous function on $[0, 1]$ and hence, by the Extreme Value Theorem (Theorem 2.16 of Chapter 13), this maximum exists. We can thus make the following definition.

Definition 5.1 *Max metric on $C[0,1]$*

The **max metric** on $C[0,1]$ is the function $d_{\max}\colon C[0,1] \times C[0,1] \to \mathbb{R}$ defined by

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [0,1]\}.$$

We show that $d_{\max}$ is a metric shortly.

Exercise 5.1

Find $d_{\max}(f,g)$ when $f, g\colon [0,1] \to \mathbb{R}$ are given by:

(a) $f(x) = 1$, $g(x) = x$ for $0 \leq x \leq 1$

(b) $f(x) = \frac{1}{2}$, $g(x) = \sin 2\pi x$ for $0 \leq x \leq 1$

(c) $f(x) = x$, $g(x) = x^2$ for $0 \leq x \leq 1$.

Hint: in answering part (c), you might find it helpful to write $h = g - f$ and to recall that the maximum and minimum of a function h can be found by considering $h(x)$ at the points x where $h'(x) = 0$ and at the endpoints $x = 0$, $x = 1$. You may also find it helpful to sketch the graph of h.

We now show that $d_{\max}$ is indeed a metric on $C[0,1]$.

Proposition 5.2

$(C[0,1], d_{\max})$ is a metric space.

Proof We show that $d_{\max}$ satisfies conditions (M1)–(M3).

(M1) Since $|g(x) - f(x)| \geq 0$ for each $x \in [0,1]$, its maximum is also non-negative. So $d_{\max}(f,g) \geq 0$ for any f and g in $C[0,1]$.

For each $f \in C[0,1]$ and each $x \in [0,1]$, we have $|f(x) - f(x)| = 0$; therefore $d_{\max}(f,f) = 0$ whenever $f \in C[0,1]$.

If f and g in $C[0,1]$ are such that $d_{\max}(f,g) = 0$ then, for each $x \in [0,1]$, we have

$$0 \leq |g(x) - f(x)| \leq d_{\max}(f,g) = 0;$$

hence $|g(x) - f(x)| = 0$ for each $x \in [0,1]$. Thus, $f(x) = g(x)$ for each $x \in [0,1]$, which implies that $f = g$.

So $d_{\max}$ satisfies (M1).

(M2) For any $f, g \in C[0,1]$ and each $x \in [0,1]$, property (M2) for the modulus function tells us that

$$|f(x) - g(x)| = |g(x) - f(x)|,$$

and so

$$d_{\max}(g, f) = \max\{|f(x) - g(x)| : x \in [0, 1]\}$$
$$= \max\{|g(x) - f(x)| : x \in [0, 1]\} = d_{\max}(f, g).$$

Thus $d_{\max}$ satisfies (M2).

(M3) Let $f, g, h \in C[0, 1]$ and $x \in [0, 1]$. Since the modulus function has property (M3), we have

$$|h(x) - f(x)| \leq |h(x) - g(x)| + |g(x) - f(x)|. \quad (1)$$

But for each $x \in [0, 1]$,

$$|h(x) - g(x)| \leq \max\{|h(x) - g(x)| : x \in [0, 1]\}$$
$$= d_{\max}(g, h), \quad (2)$$

and, similarly,

$$|g(x) - f(x)| \leq d_{\max}(f, g). \quad (3)$$

Thus for each $x \in [0, 1]$, by combining equations (1), (2) and (3), we see that

$$|h(x) - f(x)| \leq d_{\max}(f, g) + d_{\max}(g, h).$$

Since this is true for each $|h(x) - f(x)|$ with $x \in [0, 1]$, it is true for the maximum value, and so

$$d_{\max}(f, h) \leq d_{\max}(f, g) + d_{\max}(g, h).$$

Thus $d_{\max}$ satisfies (M3).

Since $d_{\max}$ satisfies conditions (M1)–(M3), it is a metric on $C[0, 1]$. Hence $(C[0, 1], d_{\max})$ is a metric space. ■

In a similar way, for any interval $[a, b]$, we can define a metric $d_{\max}$ on $C[a, b]$ by

$$d_{\max}(f, g) = \max\{|g(x) - f(x)| : x \in [a, b]\}.$$

Exercise 5.2

For $f, g \in C[0, 1]$, define $d_{\min}(f, g)$ by

$$d_{\min}(f, g) = \min\{|g(x) - f(x)| : x \in [0, 1]\}.$$

Determine whether $d_{\min}$ is a metric on $C[0, 1]$.

The next worked exercise gives another important example of a metric for $C[0, 1]$.

Worked Exercise 5.3 *Integral metric*

This worked exercise investigates a metric on the set $C[0,1]$ that is different from $d_{\max}$.

Define $d\colon C[0,1] \times C[0,1] \to \mathbb{R}$ by

$$d(f,g) = \int_0^1 |g(x) - f(x)|\,dx.$$

(a) Find $d(f,g)$ when $f, g\colon [0,1] \to \mathbb{R}$ are given by

(i) $f(x) = x$, $g(x) = 0$; (ii) $f(x) = x$, $g(x) = x^2$;
(iii) $f(x) = \sin 2\pi x$, $g(x) = 0$.

(b) Prove that d defines a metric on $C[0,1]$.

Solution

(a) (i) Here $|g(x) - f(x)| = |0 - x| = x$ when $x \in [0,1]$. Hence

$$d(f,g) = \int_0^1 x\,dx = \left[\tfrac{1}{2}x^2\right]_0^1 = \tfrac{1}{2}.$$

(ii) Here $|g(x) - f(x)| = |x^2 - x| = x - x^2$ when $x \in [0,1]$. Hence

$$d(f,g) = \int_0^1 (x - x^2)\,dx = \left[\tfrac{1}{2}x^2 - \tfrac{1}{3}x^3\right]_0^1 = \tfrac{1}{6}.$$

(iii) Here $|g(x) - f(x)| = |0 - \sin 2\pi x| = |\sin 2\pi x|$ when $x \in [0,1]$.

Now $|\sin 2\pi x| = \sin 2\pi x$ when $x \in [0, \frac{1}{2}]$, and
$|\sin 2\pi x| = -\sin 2\pi x$ when $x \in (\frac{1}{2}, 1]$. Hence

$$\begin{aligned} d(f,g) &= \int_0^{\frac{1}{2}} \sin 2\pi x\,dx - \int_{\frac{1}{2}}^1 \sin 2\pi x\,dx \\ &= \left[\frac{-1}{2\pi}\cos 2\pi x\right]_0^{\frac{1}{2}} - \left[\frac{-1}{2\pi}\cos 2\pi x\right]_{\frac{1}{2}}^1 = \frac{2}{\pi}. \end{aligned}$$

(b) To show that d is a metric on $C[0,1]$, we must verify that d satisfies (M1)–(M3).

(M1) Let $f, g \in C[0,1]$. Since $|g(x) - f(x)| \geq 0$ whenever $x \in [0,1]$, the monotonicity property of integration implies that

$$d(f,g) = \int_0^1 |g(x) - f(x)|\,dx \geq \int_0^1 0\,dx = 0.$$

Basic properties of integrals are summarised in the *Handbook*.

Direct calculation gives

$$d(f,f) = \int_0^1 |f(x) - f(x)|\,dx = \int_0^1 0\,dx = 0.$$

If $f, g \in C[0,1]$ satisfy $d(f,g) = 0$, then

$$\int_0^1 |g(x) - f(x)|\,dx = 0,$$

See the *Handbook*.

and so, since $x \mapsto |g(x) - f(x)|$ is a non-negative continuous function on $[0,1]$, it follows that $|g(x) - f(x)| = 0$ for each $x \in [0,1]$. Hence $f(x) = g(x)$ for each $x \in [0,1]$, and so $f = g$.

Thus d satisfies (M1).

(M2) Let $f, g \in C[0,1]$. Then

$$\begin{aligned} d(f,g) &= \int_0^1 |g(x) - f(x)|\,dx \\ &= \int_0^1 |f(x) - g(x)|\,dx = d(g,f), \end{aligned}$$

and so d satisfies (M2).

(M3) Let $f, g, h \in C[0,1]$. Using the Triangle Inequality for the modulus function together with the monotonicity property of integration, we have

$$\begin{aligned} d(f,h) &= \int_0^1 |h(x) - f(x)|\,dx \\ &= \int_0^1 |(h(x) - g(x)) + (g(x) - f(x))|\,dx \\ &\leq \int_0^1 (|h(x) - g(x)| + |g(x) - f(x)|)\,dx \\ &= \int_0^1 |h(x) - g(x)|\,dx + \int_0^1 |g(x) - f(x)|\,dx \\ &= d(f,g) + d(g,h). \end{aligned}$$

So d satisfies (M3).

We conclude that d is a metric on $C[0,1]$.

Open and closed balls in the max metric

Given a function $f \in C[0,1]$ and $r > 0$, the open ball $B_{d_{\max}}(f,r)$ is determined by finding the set of all functions $g \in C[0,1]$ satisfying the inequality $d_{\max}(f,g) < r$ – that is,

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [0,1]\} < r.$$

For each $x \in [0,1]$, if $|g(x) - f(x)| < r$ then $g(x)$ must lie strictly within a distance r of $f(x)$. And so, if $g \in B_{d_{\max}}(f,r)$ then for each $x \in [0,1]$, the value $g(x)$ must lie within a distance r of $f(x)$.

Thus, over the whole of $[0,1]$, $g(x)$ must lie within a band following the shape of f and extending r above and below it at each point, as Figure 5.2 illustrates. It follows that the open ball $B_{d_{\max}}(f,r)$ consists of all those $g \in C[0,1]$ whose graphs lie within this band.

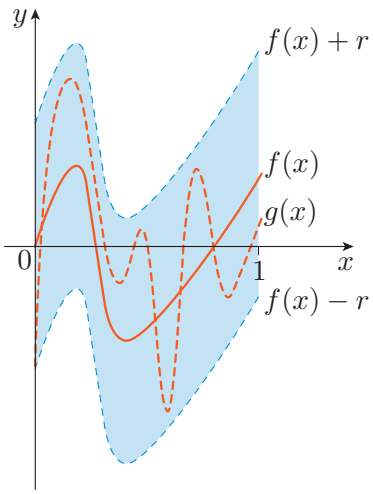


Figure 5.2 The open ball $B_{d_{\max}}(f,r)$ consists of all continuous functions whose graphs lie within the open band illustrated

The closed ball $B_{d_{\max}}[f,r]$ is obtained in a similar way except that the band includes the points exactly distance r above and below the graph of f.

Continuity and the max metric on $C[0,1]$

We now look at some functions defined on $C[0,1]$ and investigate whether they are continuous. Such functions will be a mapping from $C[0,1]$ to another metric space (Y,d). If $(Y,d)=(C[0,1],d_{\max})$, then such a mapping will map continuous functions to continuous functions. A simple example of such a mapping from $C[0,1]$ to $C[0,1]$ is given as follows.

We use the word 'map' to help us distinguish between functions *in* $C[0,1]$ and functions defined *on* $C[0,1]$.

Let us add a fixed function h from $C[0,1]$ to a function $f\in C[0,1]$; then the new function, given by $(f+h)(x)=f(x)+h(x)$ for each $x\in[0,1]$, is also continuous (being the sum of two continuous functions).

Thus we may define a map $F_h\colon C[0,1]\to C[0,1]$ by $f\mapsto(f+h)$ for $f\in C[0,1]$.

Note the use of upper case F to define this map: this helps us to distinguish it from members of $C[0,1]$.

So for example, if $h\in C[0,1]$ is given by $h(x)=x^2$, then we define the map F_h by $f\mapsto f+h$ for $f\in C[0,1]$. Thus, for each $f\in C[0,1]$, the function $F_h(f)$ is a function on $[0,1]$ whose value at a particular point $x\in[0,1]$ is given by

$$F_h(f)(x)=f(x)+x^2.$$

This is shown in Figure 5.3. Since f and h are in $C[0,1]$, $f+h$ is also in $C[0,1]$ and so $F_h\colon C[0,1]\to C[0,1]$.

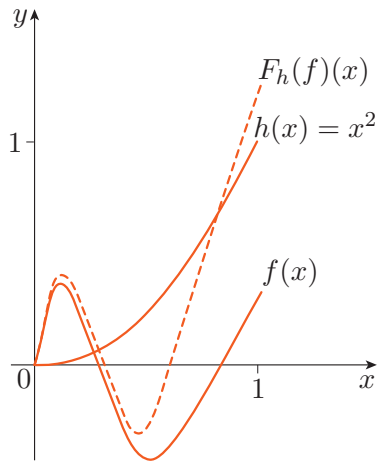


Figure 5.3 Graphs of $f(x)$, $h(x)$ and $F_h(f)(x)$

Worked Exercise 5.4

Show that for each $h\in C[0,1]$, $F_h\colon C[0,1]\to C[0,1]$ is $(d_{\max},d_{\max})$-continuous on $C[0,1]$.

Solution

Fix $h\in C[0,1]$. We show that F_h is $(d_{\max},d_{\max})$-Lipschitz and so, by Theorem 3.2, is $(d_{\max},d_{\max})$-continuous on $C[0,1]$.

Let $f,g\in C[0,1]$ – we must find an upper bound for $d_{\max}(F_h(f),F_h(g))$ in terms of $d_{\max}(f,g)$.

By definition,

$$d_{\max}(F_h(f),F_h(g))=\max\{|F_h(f)(x)-F_h(g)(x)|:x\in[0,1]\}.$$

We consider the value of $|F_h(f)(x)-F_h(g)(x)|$ for a *particular* $x\in[0,1]$. We find that

$$\begin{aligned}|F_h(f)(x)-F_h(g)(x)|&=|(f(x)+h(x))-(g(x)+h(x))|\\&=|f(x)-g(x)|.\end{aligned}$$

Hence, for each $x\in[0,1]$,

$$|F_h(f)(x)-F_h(g)(x)|=|f(x)-g(x)|$$

and so

$$\max\{|F_h(f)(x) - F_h(g)(x)| : x \in [0,1]\} = \max\{|f(x) - g(x)| : x \in [0,1]\}.$$

That is,

$$d_{\max}(F_h(g), F_h(f)) = d_{\max}(g, f).$$

Thus F_h is $(d_{\max}, d_{\max})$-Lipschitz (with Lipschitz constant 1) and the result follows.

We have seen that adding a fixed continuous function to a function in $C[0,1]$ produces a continuous mapping from $(C[0,1], d_{\max})$ to itself. Similarly, multiplying by a fixed continuous function also produces a continuous mapping from $(C[0,1], d_{\max})$ to itself: we choose a function $h \in C[0,1]$ and define the *product* function by $G_h(f) = hf$, for $f \in C[0,1]$. Since f and h are in $C[0,1]$, the product hf is also in $C[0,1]$ and so $G_h : C[0,1] \to C[0,1]$.

Thus $G_h(f)(x) = h(x) \times f(x)$.

Worked Exercise 5.5

Show that for each $h \in C[0,1]$, G_h is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$.

Solution

The proof is very similar to that of Worked Exercise 5.4.

Fix $h \in C[0,1]$.

We show that G_h is $(d_{\max}, d_{\max})$-Lipschitz and so, by Theorem 3.2, is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$. Let $f, g \in C[0,1]$ – we must find an upper bound for $d_{\max}(G_h(f), G_h(g))$ in terms of $d_{\max}(f, g)$.

As in Worked Exercise 5.4, we know that

$$d_{\max}(G_h(f), G_h(g)) = \max\{|G_h(f)(x) - G_h(g)(x)| : x \in [0,1]\}.$$

We consider the value of $|G_h(f)(x) - G_h(g)(x)|$ for a *particular* $x \in [0,1]$. We find that

$$\begin{aligned} |G_h(f)(x) - G_h(g)(x)| &= |h(x)f(x) - h(x)g(x)| \\ &= |h(x)(f(x) - g(x))| \\ &= |h(x)|\,|f(x) - g(x)| \leq |h(x)|\, d_{\max}(g, f). \end{aligned}$$

So we must find an upper bound for $|h(x)|$ in which x does not appear. But $h \in C[0,1]$, and so $|h| \in C[0,1]$. Hence, by the Boundedness Theorem, there is a real number $M \geq 0$ for which

$$|h(x)| \leq M \text{ for each } x \in [0,1].$$

This is Theorem 2.15 in Chapter 13.

It follows that for each $x \in [0,1]$,

$$|h(x)f(x) - h(x)g(x)| \leq M d_{\max}(f,g)$$

and so

$$d_{\max}(G_h(f), G_h(g)) \leq M d_{\max}(f,g).$$

Thus G_h is $(d_{\max}, d_{\max})$-Lipschitz (with Lipschitz constant M, the bound on h) and the result follows.

We now look at a few examples where our mappings go from $C[0,1]$ to a different metric space, $(\mathbb{R}, d^{(1)})$. We again make frequent use of Theorem 3.2 and show that our mappings are Lipschitz in order to show that they are continuous.

Worked Exercise 5.6

Show that $I\colon C[0,1] \to \mathbb{R}$ defined by $I(f) = \int_0^1 f(x)\,dx$ is $(d_{\max}, d^{(1)})$-continuous.

Solution

We show that I is a $(d_{\max}, d^{(1)})$-Lipschitz function and so, by Theorem 3.2, it is $(d_{\max}, d^{(1)})$-continuous.

Let $f, g \in C[0,1]$ be given. We aim to find an upper bound for $d^{(1)}(I(f), I(g))$ in terms of $d_{\max}(f,g)$. Now

$$\begin{aligned} d^{(1)}(I(f), I(g)) &= \left| \int_0^1 g(x)\,dx - \int_0^1 f(x)\,dx \right|, \text{ using definition of } I \\ &= \left| \int_0^1 (g(x) - f(x))\,dx \right|. \end{aligned} \tag{4}$$

The definition of $d_{\max}$ implies that for each $x \in [0,1]$,

$$|g(x) - f(x)| \leq d_{\max}(f,g),$$

which means that for each $x \in [0,1]$,

$$-d_{\max}(f,g) \leq g(x) - f(x) \leq d_{\max}(f,g).$$

Integrating the terms in this double inequality between 0 and 1, we obtain

$$\int_0^1 -d_{\max}(f,g)\,dx \leq \int_0^1 (g(x) - f(x))\,dx \leq \int_0^1 d_{\max}(f,g)\,dx;$$

that is,

$$-d_{\max}(f,g) \leq \int_0^1 (g(x) - f(x))\,dx \leq d_{\max}(f,g),$$

Note that $\int_0^1 d_{\max}(f,g)\,dx$ is just the integral of a constant between 0 and 1 and so takes the value of the constant.

which is equivalent to

$$\left| \int_0^1 (g(x) - f(x))\, dx \right| \leq d_{\max}(f, g).$$

This, together with equation (4), gives

$$d^{(1)}(I(f), I(g)) \leq d_{\max}(f, g).$$

Hence I is $(d_{\max}, d^{(1)})$-Lipschitz (with Lipschitz constant 1), and so by Theorem 3.2 I is $(d_{\max}, d^{(1)})$-continuous on $C[0, 1]$.

Exercise 5.3

Let $h \in C[0, 1]$ be fixed. Define $I_h \colon C[0, 1] \to \mathbb{R}$ by

$$I_h(f) = \int_0^1 h(x) f(x)\, dx.$$

Show that I_h is $(d_{\max}, d^{(1)})$-continuous on $C[0, 1]$.

Hint: you may find Proposition 4.4 of Chapter 14, the Composition Rule, useful.

Worked Exercise 5.7

For $f \in C[0, 1]$, define a function $F(f) \colon [0, 1] \to \mathbb{R}$ by

$$F(f)(t) = \int_0^t f(x)\, dx.$$

(a) Show that $F(f) \in C[0, 1]$.

(b) Show that the function $F \colon C[0, 1] \to C[0, 1]$ is $(d_{\max}, d_{\max})$-continuous.

Hint: show that $|F(f)(t) - F(f)(s)| \leq |t - s|\, d_{\max}(f, 0)$ for each $s, t \in [0, 1]$.

Solution

(a) Let $f \in C[0, 1]$ and $s, t \in [0, 1]$ with $s \leq t$. Then

$$\begin{aligned} |F(f)(t) - F(f)(s)| &= \left| \int_0^t f(x)\, dx - \int_0^s f(x)\, dx \right| \\ &= \left| \int_s^t f(x)\, dx \right| \leq \int_s^t |f(x)|\, dx \\ &\leq \int_s^t d_{\max}(f, 0)\, dx \\ &= (t - s) d_{\max}(f, 0) = |t - s|\, d_{\max}(f, 0). \end{aligned}$$

Hence $F(f)$ is a $(d^{(1)}, d^{(1)})$-Lipschitz function on $[0,1]$ (with Lipschitz constant $d_{\max}(f,0)$), and so $F(f)$ is $(d^{(1)}, d^{(1)})$-continuous by Theorem 3.2; that is, $F(f) \in C[0,1]$.

(b) In order to show that F is $(d_{\max}, d_{\max})$-continuous, let $f, g \in C[0,1]$, and observe that for each $t \in [0,1]$,

$$\begin{aligned} |F(f)(t) - F(g)(t)| &= \left| \int_0^t (f(x) - g(x))\, dx \right| \\ &\leq \int_0^t |f(x) - g(x)|\, dx \\ &\leq \int_0^1 d_{\max}(f,g)\, dx = d_{\max}(f,g). \end{aligned}$$

Hence

$$d_{\max}(F(f), F(g)) \leq d_{\max}(f,g);$$

thus F is a $(d_{\max}, d_{\max})$-Lipschitz function and so, in particular, is a $(d_{\max}, d_{\max})$-continuous function.

5.1 Sequences in $C[0,1]$

We finish this chapter by looking at the relationship of $d_{\max}$-convergence of sequences in $C[0,1]$ to pointwise and uniform convergence.

Recall that $f \in C[0,1]$ if $f\colon [0,1] \to \mathbb{R}$ and f is continuous on $[0,1]$ for the Euclidean metric. The distance $d_{\max}$ between two functions f and g in $C[0,1]$ is defined by

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [0,1]\}$$

(we saw in Proposition 5.2 that this defines a metric on $C[0,1]$).

Definition 3.2 of Chapter 14 tells us that a sequence of functions (f_n) in $C[0,1]$ converges to f in $C[0,1]$ for $d_{\max}$ if, and only if, $d_{\max}(f_n, f) \to 0$ as $n \to \infty$; that is if, and only if,

$$\max\{|f(x) - f_n(x)| : x \in [0,1]\} \to 0 \text{ as } n \to \infty.$$

The resemblance between this and the definition of uniform convergence means that the following result is not a great surprise.

Theorem 5.8 *Uniform Convergence Theorem*

Let (f_n) be a sequence of functions in $C[0,1]$.

If (f_n) converges uniformly on $[0,1]$ to the function $f\colon [0,1] \to \mathbb{R}$, then f is continuous on $[0,1]$ (and so is in $C[0,1]$) and $d_{\max}(f_n, f) \to 0$ as $n \to \infty$.

Conversely, if $f \in C[0,1]$ and $d_{\max}(f_n, f) \to 0$ as $n \to \infty$, then (f_n) converges uniformly to f on $[0,1]$.

Proof Let (f_n) be a sequence in $C[0,1]$ that converges uniformly to the function $f\colon [0,1] \to \mathbb{R}$. By Theorem 4.7 of this chapter, f is continuous on $[0,1]$. It remains to show that $(d_{\max}(f_n, f))_{n\in\mathbb{N}}$ is a null sequence.

Since $f_n \to f$ uniformly on $[0,1]$, we know (M_n) is a real null sequence when M_n is given by

$$M_n = \sup\{|f_n(x) - f(x)| : x \in [0,1]\}.$$

In addition, since $[0,1]$ is a closed bounded interval, the Extreme Value Theorem (Theorem 2.16 of Chapter 13) tells us that $f_n - f$ attains its bounds for each $n \in \mathbb{N}$. Thus

$$M_n = \max\{|f_n(x) - f(x)| : x \in [0,1]\} = d_{\max}(f_n, f).$$

Hence $(d_{\max}(f_n, f))$ is a null sequence, as claimed.

Conversely, let (f_n) be a sequence of functions in $C[0,1]$ that converges to $f \in C[0,1]$ with respect to $d_{\max}$. Then $(d_{\max}(f_n, f))$ is a null sequence. Again we can deduce from the Extreme Value Theorem that $f_n - f$ is bounded for each $n \in \mathbb{N}$ and that for each $n \in \mathbb{N}$,

$$\begin{aligned} M_n &= \sup\{|f_n(x) - f(x)| : x \in [0,1]\} \\ &= \max\{|f_n(x) - f(x)| : x \in [0,1]\} = d_{\max}(f_n, f), \end{aligned}$$

so (M_n) is also a null sequence. Hence $f_n \to f$ uniformly as $n \to \infty$. ■

Thus, if we are given a sequence of functions (f_n) in $C[0,1]$, we can use our knowledge of the properties of uniform convergence to determine whether it is convergent for $d_{\max}$, as follows.

Strategy: to determine convergence of sequences of functions in $C[0,1]$

1. Identify the possible limit function $f \in C[0,1]$ by finding the pointwise limit of (f_n).

   If there is no pointwise limit, then Lemma 4.5 implies that there is no uniform limit, and so Theorem 5.8 tells us that (f_n) is not convergent with respect to $d_{\max}$.

2. If the pointwise limit of (f_n) exists, determine whether it is continuous.

   If it is not continuous, then Lemma 4.5 tells us that the only possible uniform limit is not continuous, and Theorems 4.7 and 5.8 together then tell us that (f_n) is not convergent with respect to $d_{\max}$.

3. If f is the pointwise limit of (f_n) and is continuous, determine whether $(d_{\max}(f_n, f))$ is a null sequence.

   If it is, then by definition the sequence (f_n) converges to f for $d_{\max}$. Otherwise it does not converge in $C[0,1]$ for $d_{\max}$.

Worked Exercise 5.9

Consider the sequence of functions (f_n) where for each $n \in \mathbb{N}$, $f_n\colon [0,1] \to \mathbb{R}$ is given by

$$f_n(x) = \begin{cases} 1 - nx, & \text{if } 0 \leq x < \frac{1}{n}, \\ 0, & \text{if } \frac{1}{n} \leq x \leq 1. \end{cases}$$

(a) For each $x \in [0,1]$, find $\lim_{n\to\infty} f_n(x)$.

(b) Use the strategy to determine whether the sequence (f_n) converges in $C[0,1]$ for $d_{\max}$.

Solution

(a) Sketches of the graphs of f_n for the first few values of n (Figure 5.4) suggest that for each $x > 0$, $f_n(x)$ is eventually 0, whereas if $x = 0$, then $f_n(x) = 1$ for each $n \in \mathbb{N}$. Indeed, if $0 < x \leq 1$, then $f_n(x) = 0$ for $n > \frac{1}{x}$, and hence $f_n(x) \to 0$ as $n \to \infty$.

In addition $f_n(0) = 1$ for each $n \in \mathbb{N}$, and so $f_n(0) \to 1$ as $n \to \infty$.

Hence the pointwise limit of the sequence (f_n) is the function $f\colon [0,1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{if } 0 < x \leq 1. \end{cases}$$

(b) We use the strategy given above. Step 1 requires us to identify a possible limit function. By part (a), the only possible limit function for the sequence (f_n) is the pointwise limit function $f\colon [0,1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{if } 0 < x \leq 1. \end{cases}$$

But this function f is *not* continuous on $[0,1]$, and so is not in $C[0,1]$. Hence (f_n) does not converge in $C[0,1]$ for $d_{\max}$.

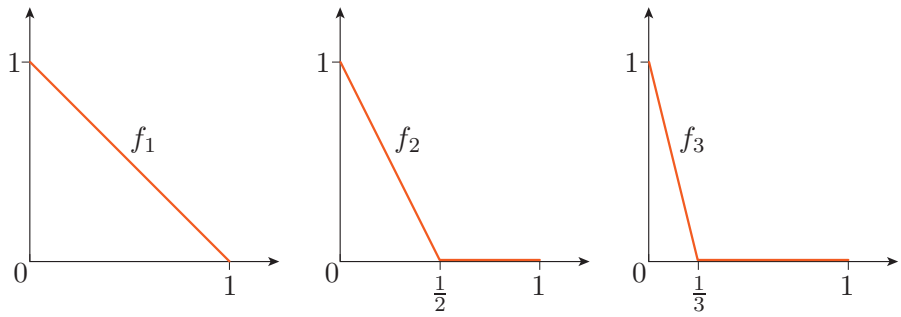


Figure 5.4 Sketches of f_1, f_2 and f_3 for Worked Exercise 5.9

Exercise 5.4

Consider the sequence of functions (f_n) where for each $n \in \mathbb{N}$, $f_n \colon [0,1] \to \mathbb{R}$ is given by

$$f_n(x) = \begin{cases} 2nx, & \text{if } 0 \leq x < \frac{1}{2n}, \\ 2 - 2nx, & \text{if } \frac{1}{2n} \leq x < \frac{1}{n}, \\ 0, & \text{if } \frac{1}{n} \leq x \leq 1. \end{cases}$$

(a) Sketch the graphs of f_1, f_2 and f_3.

(b) For each $x \in [0,1]$, find $\lim_{n\to\infty} f_n(x)$.

(c) Determine whether the sequence (f_n) converges in $C[0,1]$ for $d_{\max}$.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should be able to:

(a) Use a one–one map to transfer a metric from one set W to another set X.

(b) Given a metric on a set X, describe the induced metric on a subset of X.

(c) Construct a metric on the product of metric spaces.

(d) Define and use the Hamming distance.

(e) Define and use the Cantor metric.

(f) State the definition of a Lipschitz function, and determine whether a given function is Lipschitz.

(g) Explain why Lipschitz functions are continuous.

(h) Recognise that different metrics can give rise to the same continuous functions.

(i) Understand the relationship between continuity in a product space and its constituent spaces.

(j) Define pointwise convergence and uniform convergence for sequences of functions.

(k) Determine whether a given sequence of functions converges pointwise.

(l) Determine whether a given sequence of functions converges uniformly.

(m) Define and use the max metric on $C[0,1]$.

(n) Find the limit of a given sequence of functions in $C[0,1]$ for $d_{\max}$.

Solutions to exercises

Solution to Exercise 1.1

The distance function d is obtained from the pull-back of the Euclidean metric $d^{(1)}$ by the function $f(x) = \tan^{-1} x$. Hence, by Theorem 1.2, it is enough to show that f is one–one on $\mathbb{R}$ to conclude that d is a metric. Since $\tan^{-1}$ is a strictly increasing function (see Figure 1.2), it is indeed one–one. Thus d is a metric.

Solution to Exercise 1.2

The distance function d is obtained from the pull-back of the Euclidean metric $d^{(1)}$ by the function $f(x) = x^2$. However, f is not one–one, and so, by the remark after Theorem 1.2, d is not a metric.

(You can also check this directly since, for each $x \in \mathbb{R}$, $d(x, -x) = 0$, violating condition (M1) when $x \neq 0$.)

Solution to Exercise 1.3

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$ is given. Let $A \subseteq X$ and let d_A be the metric induced on A by d.

Suppose that $a \in A$ and f is (d, e)-continuous at a.

Let $(a_n)_{n\in\mathbb{N}}$ be a sequence in A that d_A-converges to a. (Thus $(d_A(a_n, a))$ is a null sequence.)

Then, from Theorem 1.5, (a_n) is a sequence in X that d-converges to a and so, since f is (d, e)-continuous at a, we conclude that the sequence $(f(a_n))$ e-converges to $f(a)$. But $f(a) = f|_A(a)$ and for each $n \in \mathbb{N}$, $f(a_n) = f|_A(a_n)$. Hence $(f|_A(a_n))$ e-converges to $f|_A(a)$ as $n \to \infty$.

In other words, for any choice of sequence (a_n) in A that d_A-converges to a, we have $(f|_A(a_n))$ e-converges to $f|_A(a)$ and so $f|_A$ is (d_A, e)-continuous at a.

Solution to Exercise 1.4

We find that

$$d^{(3)}(\mathbf{0}, (1, 0, 0)) = \sqrt{(1-0)^2 + (0-0)^2 + (0-0)^2} = 1,$$
$$d^{(3)}(\mathbf{0}, (0, 0, 1)) = \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2} = 1,$$

and so $(1, 0, 0)$ and $(0, 0, 1)$ are in S. Finally,

$$d_S^{(3)}((1, 0, 0), (0, 0, 1)) = \sqrt{(0-1)^2 + (0-0)^2 + (1-0)^2} = \sqrt{2}.$$

Solution to Exercise 1.5

(a) Let $\mathbf{a} = (a_1, a_2) \in S$ and $\mathbf{b} = (b_1, b_2) \in S$. In order to show that f is one–one, we need to show that if $f(\mathbf{a}) = f(\mathbf{b})$, then $\mathbf{a} = \mathbf{b}$.

So suppose that $f(\mathbf{a}) = f(\mathbf{b})$. Since, whatever the choice of $\mathbf{a} \in S$, $0 \leq f(\mathbf{a}) \leq 3$, we consider three cases.

$0 \leq f(\mathbf{a}) < 1$**:** In this case, $a_1 = 0$ and $f(\mathbf{a}) = 1 - a_2 = f(\mathbf{b})$ by our assumption. Now $a_2 \in [0, 1]$, and so $f(\mathbf{b}) \in [0, 1]$, which implies $b_1 = 0$ and $f(\mathbf{b}) = 1 - b_2$. That is, $1 - a_2 = 1 - b_2$ and $a_1 = 0 = b_1$, giving $\mathbf{a} = \mathbf{b}$.

$1 \leq f(\mathbf{a}) \leq 2$**:** In this case, $a_2 = 0$ and $f(\mathbf{a}) = a_1 + 1 = f(\mathbf{b})$ by our assumption. Now $f(\mathbf{a}) \in [1, 2]$, and so $f(\mathbf{b}) \in [1, 2]$, which implies $b_2 = 0$ and $f(\mathbf{b}) = b_1 + 1$. That is, $a_1 + 1 = b_1 + 1$ and $a_2 = b_2 = 0$, giving $\mathbf{a} = \mathbf{b}$.

$2 < f(\mathbf{a}) \leq 3$**:** In this case, $a_1 = 1$ and $f(\mathbf{a}) = a_2 + 2 = f(\mathbf{b})$ by our assumption. Now $a_2 + 2 \in (2, 3]$, and so $f(\mathbf{b}) \in (2, 3]$, which implies $b_1 = 1$ and $f(\mathbf{b}) = b_2 + 2$. That is, $a_2 + 2 = b_2 + 2$ and $a_1 = 1 = b_1$, giving $\mathbf{a} = \mathbf{b}$.

Thus in all three cases, $f(\mathbf{a}) = f(\mathbf{b})$ implies $\mathbf{a} = \mathbf{b}$. We conclude that f is one–one.

(b) Since f is a one–one mapping from S into $\mathbb{R}$ and d is the pull-back of the Euclidean metric on $\mathbb{R}$ by f, $d_f^{(1)}$, we deduce from Theorem 1.2 that d is a metric on S.

(c) We see that $f(\mathbf{a}) = f((0, 1)) = 0$ and $f(\mathbf{b}) = f((1, 1)) = 3$, so $d(\mathbf{a}, \mathbf{b}) = |0 - 3| = 3$. (Alternatively, since d measures distance 'along' S and S has length 3, it follows that $d(\mathbf{a}, \mathbf{b}) = 3$.)

In contrast, the Euclidean distance between the points $\mathbf{a}$ and $\mathbf{b}$ is 1 and so $d_S^{(2)}(\mathbf{a}, \mathbf{b}) = 1$.

Solution to Exercise 1.6

The proof that e_1 is a metric closely follows the proof given for the taxicab metric in Chapter 14.

(M1) For each $\mathbf{x}$, $\mathbf{y} \in X_1 \times X_2$, $d_1(x_1, y_1)$ and $d_2(x_2, y_2)$ are non-negative, since d_1 and d_2 are metrics, and hence so is their sum: thus $e_1(\mathbf{x}, \mathbf{y}) \geq 0$.

For each $\mathbf{x} \in X_1 \times X_2$,

$$e_1(\mathbf{x}, \mathbf{x}) = d_1(x_1, x_1) + d_2(x_2, x_2) = 0 + 0 = 0,$$

since d_1 and d_2 are metrics.

Conversely, suppose $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$ are such that $e_1(\mathbf{x}, \mathbf{y}) = 0$. Then

$$0 = d_1(x_1, y_1) + d_2(x_2, y_2).$$

Since d_1 and d_2 are metrics, $d_1(x_1, y_1) \geq 0$ and $d_2(x_2, y_2) \geq 0$, so this equation implies that $d_1(x_1, y_1) = 0$ and $d_2(x_2, y_2) = 0$. Thus, since d_1 and d_2 are metrics, $x_1 = y_1$ and $x_2 = y_2$; that is, $\mathbf{x} = \mathbf{y}$. So e_1 satisfies (M1).

(M2) Let $\mathbf{x}, \mathbf{y} \in X_1 \times X_2$. Using property (M2) for the metrics d_1 and d_2, we have

$$\begin{aligned} e_1(\mathbf{y}, \mathbf{x}) &= d_1(y_1, x_1) + d_2(y_2, x_2) \\ &= d_1(x_1, y_1) + d_2(x_2, y_2) = e_1(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus e_1 satisfies (M2).

(M3) Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in X_1 \times X_2$. Then, using property (M3) for the metrics d_1 and d_2,

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{z}) &= d_1(x_1, z_1) + d_2(x_2, z_2) \\ &\leq (d_1(x_1, y_1) + d_1(y_1, z_1)) + (d_2(x_2, y_2) + d_2(y_2, z_2)) \\ &= (d_1(x_1, y_1) + d_2(x_2, y_2)) + (d_1(y_1, z_1) + d_2(y_2, z_2)) \\ &= e_1(\mathbf{x}, \mathbf{y}) + e_1(\mathbf{y}, \mathbf{z}). \end{aligned}$$

Thus e_1 satisfies (M3).

Therefore e_1 is a metric on $X_1 \times X_2$.

Solution to Exercise 1.7

$$\begin{aligned} e_1(\mathbf{0}, (2,1)) &= |2-0| + |1-0| = 3, \\ e_2(\mathbf{0}, (2,1)) &= \sqrt{(2-0)^2 + (1-0)^2} = \sqrt{5}, \\ e_\infty(\mathbf{0}, (2,1)) &= \max\{|2-0|, |1-0|\} = 2. \end{aligned}$$

Since these distances are all different, we conclude that the three metrics are distinct.

Solution to Exercise 1.8

(a) For the unit open ball for e_∞ centred at the origin, we have to find those $(x_1, x_2) \in \mathbb{R}^2$ satisfying

$$\max\{|x_1|, |x_2|\} < 1.$$

This is satisfied provided both $-1 < x_1 < 1$ and $-1 < x_2 < 1$, and the points satisfying these inequalities form the interior of the square shown in Figure 1.6.

(b) The unit closed balls centred at the origin are the regions shown in Figure 1.6 together with their boundary points, comprising the bounding diamond, circle and square.

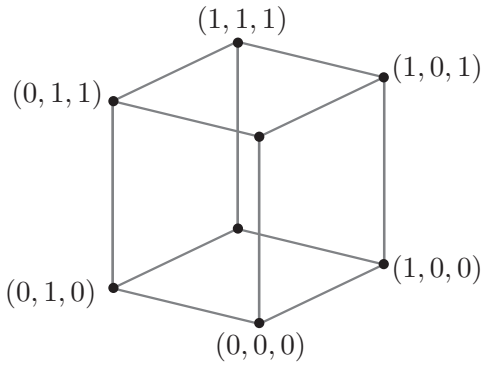
The unit spheres centred at the origin are simply the bounding diamond, circle and square of the regions shown in Figure 1.6.

Solution to Exercise 1.9

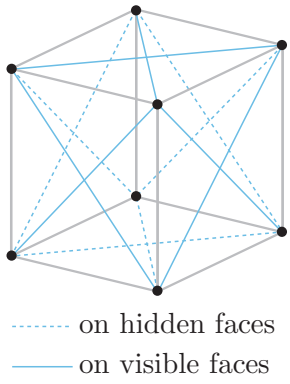
When $n = 3$,

$$A^n = A^3 = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}$$

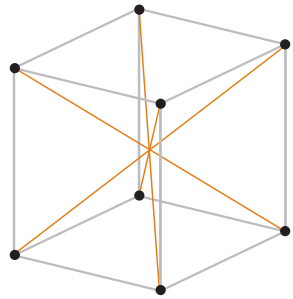
and this set of binary strings can be represented as the vertices of the usual unit cube in $\mathbb{R}^3$. Adjacent vertices are distance one apart and the distance between non-adjacent vertices is determined by the minimum number of edges required to join them, as shown in the figure.



(a) points that are distance 1 apart



(b) points that are distance 2 apart



(c) points that are distance 3 apart

Thus

$$B_{d_H}((0,0,0),r) = \begin{cases} \{(0,0,0)\}, & \text{if } 0 < r \leq 1, \\ \{(0,0,0),(0,0,1),(0,1,0),(1,0,0)\}, & \text{if } 1 < r \leq 2, \\ A^3 - \{(1,1,1)\}, & \text{if } 2 < r \leq 3, \\ A^3, & \text{if } 3 < r. \end{cases}$$

Solution to Exercise 2.1

Here $\mathbf{a}$ and $\mathbf{e}$ first differ at their second term, so $d_{\mathbf{C}}(\mathbf{a},\mathbf{e}) = 2^{-2} = \frac{1}{4}$; $\mathbf{b}$ and $\mathbf{e}$ first differ at their third term, so $d_{\mathbf{C}}(\mathbf{b},\mathbf{e}) = 2^{-3} = \frac{1}{8}$; $\mathbf{c}$ and $\mathbf{e}$ first differ at their fourth term, so $d_{\mathbf{C}}(\mathbf{c},\mathbf{e}) = 2^{-4} = \frac{1}{16}$; and $\mathbf{d}$ and $\mathbf{e}$ first differ at their fifth term, so $d_{\mathbf{C}}(\mathbf{d},\mathbf{e}) = 2^{-5} = \frac{1}{32}$.

Solution to Exercise 2.2

We show that d satisfies conditions (M1)–(M3).

(M1) Since $|x| \geq 0$ whenever $x \in \mathbb{R}$, it follows that, for $(a_n), (b_n) \in \ell_0$,

$$d((a_n), (b_n)) = \sum_{n=1}^{\infty} |b_n - a_n| \geq 0.$$

Also,

$$d((a_n), (a_n)) = \sum_{n=1}^{\infty} |a_n - a_n| = 0,$$

and if $d((a_n), (b_n)) = 0$, then $\sum_{n=1}^{\infty} |b_n - a_n| = 0$, and so $a_n = b_n$ for each $n \in \mathbb{N}$. Hence in this case $(a_n) = (b_n)$.

Thus (M1) holds.

(M2) Let $(a_n), (b_n) \in \ell_0$. Then

$$\begin{aligned} d((a_n), (b_n)) &= \sum_{n=1}^{\infty} |b_n - a_n| \\ &= \sum_{n=1}^{\infty} |a_n - b_n| = d((b_n), (a_n)), \end{aligned}$$

and so (M2) holds.

(M3) Let $(a_n), (b_n), (c_n) \in \ell_0$. Then, for any $n \in \mathbb{N}$, the Triangle Inequality for the modulus function gives

$$|c_n - a_n| \leq |c_n - b_n| + |b_n - a_n|,$$

and so, using this inequality for each term of the sum,

$$\begin{aligned} \sum_{n=1}^{\infty} |c_n - a_n| &\leq \sum_{n=1}^{\infty} (|c_n - b_n| + |b_n - a_n|) \\ &= \sum_{n=1}^{\infty} |c_n - b_n| + \sum_{n=1}^{\infty} |b_n - a_n|. \end{aligned}$$

All of these sums have only a finite number of non-zero terms, so there is no need to call upon any of the theory of convergent series.

Thus

$$d((a_n), (c_n)) \leq d((a_n), (b_n)) + d((b_n), (c_n)),$$

as required.

Hence d is a metric on ℓ_0.

Solution to Exercise 2.3

Let $\mathbf{a} = (a_1, a_2, \ldots) \in \mathbf{C}$ and fix $n \in \mathbb{N}$. Recall that by definition

$$B_d[\mathbf{a}, 2^{-n}] = \{\mathbf{b} \in \mathbf{C} : d(\mathbf{a}, \mathbf{b}) \leq 2^{-n}\}.$$

If $n = 1$, then $2^{-n} = \frac{1}{2}$ and every element of $\mathbf{C}$ is within distance $\frac{1}{2}$ of $\mathbf{a}$. So in this case $B_d[\mathbf{a}, \frac{1}{2}] = \mathbf{C}$.

Suppose now that $n \geq 2$.

If $\mathbf{b} \in \mathbf{C}$ and there is an $i \in \{1, 2, \ldots, n-1\}$ for which $a_i \neq b_i$, then

$$d(\mathbf{a}, \mathbf{b}) \geq 2^{-i} \geq 2^{-n-1} > 2^{-n}$$

and so such a $\mathbf{b}$ is not in $B_d[\mathbf{a}, 2^{-n}]$.

Conversely, if $\mathbf{b} \in \mathbf{C}$ agrees with $\mathbf{a}$ for the first $n-1$ terms, then $\mathbf{b} = (a_1, a_2, \ldots, a_{n-1}, b_n, b_{n+1}, \ldots)$ and $\mathbf{b}$ cannot differ from $\mathbf{a}$ before the nth term. In which case, $d(\mathbf{a}, \mathbf{b}) \leq 2^{-n}$, implying that $\mathbf{b} \in B_d[\mathbf{a}, 2^{-n}]$.

We conclude that in this case

$$B_d[\mathbf{a}, 2^{-n}] = \{(a_1, a_2, a_3, \ldots, a_{n-1}, c_1, c_2, c_3, \ldots) : (c_i) \in \mathbf{C}\},$$

as required.

Solution to Exercise 3.1

We show that f is $(d_H, d_{\mathbf{C}})$-Lipschitz and use Theorem 3.2.

Let $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in X. If $\mathbf{a} = \mathbf{b}$, then $d_H(\mathbf{a}, \mathbf{b}) = 0$ and $f(\mathbf{a}) = f(\mathbf{b})$ and so, in this case, we find

$$d_{\mathbf{C}}(f(\mathbf{a}), f(\mathbf{b})) = 0 = d_H(\mathbf{a}, \mathbf{b}).$$

If $\mathbf{a} \neq \mathbf{b}$, then they differ at between 1 and n places, so $1 \leq d_H(\mathbf{a}, \mathbf{b}) \leq n$. But it is always the case that $d_{\mathbf{C}}(\mathbf{x}, \mathbf{y}) \leq 2^{-1}$ no matter the choice of $\mathbf{x}$ and $\mathbf{y}$, so

$$d_{\mathbf{C}}(f(\mathbf{a}), f(\mathbf{b})) \leq 2^{-1} \leq 1 \leq d_H((a_1, a_2, a_3, \ldots, a_n), (b_1, b_2, b_3, \ldots, b_n)).$$

Hence f is $(d_H, d_{\mathbf{C}})$-Lipschitz (with Lipschitz constant 1) and we conclude by Theorem 3.2 that f is $(d_H, d_{\mathbf{C}})$-continuous on A^n.

Solution to Exercise 3.2

Let $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$. Then

$$\begin{aligned} e(\mathbf{x}, \mathbf{y}) &= |y_1 - x_1| + 2|y_2 - x_2| \\ &\leq 2|y_1 - x_1| + 2|y_2 - x_2| \\ &= 2(|y_1 - x_1| + |y_2 - x_2|) = 2e_1(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Similarly,

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= |y_1 - x_1| + |y_2 - x_2| \\ &\leq |y_1 - x_1| + 2|y_2 - x_2| \\ &= e(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Thus,

$$e_1(\mathbf{x}, \mathbf{y}) \leq e(\mathbf{x}, \mathbf{y}) \leq 2e_1(\mathbf{x}, \mathbf{y}).$$

Hence e and e_1 are equivalent.

Solution to Exercise 3.3

(a) Fix $f\colon X \to Y$ and $x \in X$.

Recall that f is (d_1, e)-continuous at x if, and only if, whenever (x_n) is a sequence in X for which $x_n \xrightarrow{d_1} x$ as $n \to \infty$, then $f(x_n) \xrightarrow{e} f(x)$ as $n \to \infty$.

But Proposition 3.7 tells us that $x_n \xrightarrow{d_1} x$ as $n \to \infty$ if, and only if, $x_n \xrightarrow{d_2} x$ as $n \to \infty$.

Hence f is (d_1, e)-continuous at x if, and only if, whenever (x_n) is a sequence in X for which $x_n \xrightarrow{d_2} x$ as $n \to \infty$, then $f(x_n) \xrightarrow{e} f(x)$ as $n \to \infty$, which is the same as f being (d_2, e)-continuous at x.

(b) Fix $g\colon Y \to X$ and $y \in Y$.

Recall that g is (e, d_1)-continuous at y if, and only if, whenever (y_n) is a sequence in Y for which $y_n \xrightarrow{e} y$ as $n \to \infty$, then $g(y_n) \xrightarrow{d_1} g(y)$ as $n \to \infty$.

But Proposition 3.7 tells us that for any sequence (x_n) in X, $x_n \xrightarrow{d_1} x$ as $n \to \infty$ if, and only if, $x_n \xrightarrow{d_2} x$ as $n \to \infty$, and so $g(y_n) \xrightarrow{d_1} g(y)$ if, and only if, $g(y_n) \xrightarrow{d_2} g(y)$ as $n \to \infty$.

Hence g is (e, d_1)-continuous at y if, and only if, whenever (y_n) is a sequence in Y for which $y_n \xrightarrow{e} y$ as $n \to \infty$, then $g(y_n) \xrightarrow{d_2} g(y)$ as $n \to \infty$, which is the same as g being (e, d_2)-continuous at y.

Solution to Exercise 3.4

(a) (i) $(0, (0, 0, 0, 0, 0, 0, \ldots))$.
(ii) $(0, (1, 0, 1, 0, 1, 0, \ldots))$.
(Notice the extra brackets.)

(b) Consider the composites of the projection functions with f, that is, $p_1 \circ f\colon \mathbf{C} \to \mathbb{R}$ and $p_2 \circ f\colon \mathbf{C} \to \mathbf{C}$. Now

$$(p_1 \circ f)(\mathbf{a}) = a_1 = g(\mathbf{a}),$$

where $g\colon \mathbf{C} \to \mathbb{R}$ is given by $g(\mathbf{a}) = a_1$. By assumption, g is $(d_{\mathbf{C}}, d^{(1)})$-continuous. Also,

$$(p_2 \circ f)(\mathbf{a}) = (a_2, a_3, a_4, \ldots) = \sigma(\mathbf{a}),$$

and we saw in Worked Exercise 2.6 that σ is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous.

Since both $p_1 \circ f$ and $p_2 \circ f$ are continuous, it follows from Theorem 3.12 that f is $(d_{\mathbf{C}}, e)$-continuous, where e is any one of e_1, e_2 or e_∞.

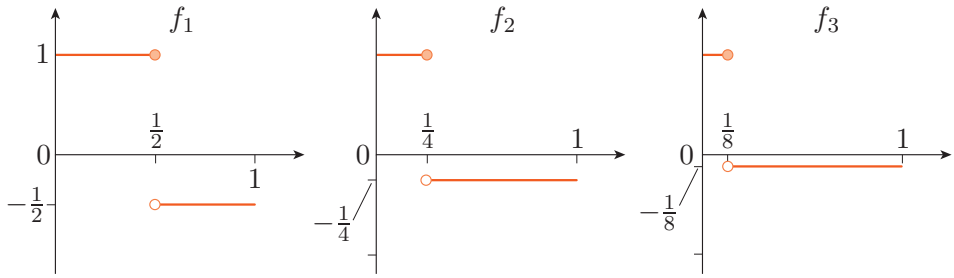
Solution to Exercise 3.5

The first component sequence converges for the Euclidean metric, since it is a basic null sequence.

However, the sequence $(\frac{1}{n^2})$ is not eventually constant and so, by the result of Exercise 3.2 of Chapter 14, does not converge for the discrete metric. Hence $((\frac{1}{n}, \frac{1}{n^2}))$ is divergent for this metric.

Solution to Exercise 4.1

Sketching the graphs of the functions f_n for $n = 1, 2$ and 3 suggests that, for each $x > 0$, $f_n(x) \to 0$ as $n \to \infty$.



Indeed, for $x > 0$, choose $N \in \mathbb{N}$ so that $\frac{1}{2^N} < x$; then, for each $n > N$, $f_n(x) = -\frac{1}{2^n}$. So, for $x > 0$, $f_n(x) \to 0$ as $n \to \infty$.

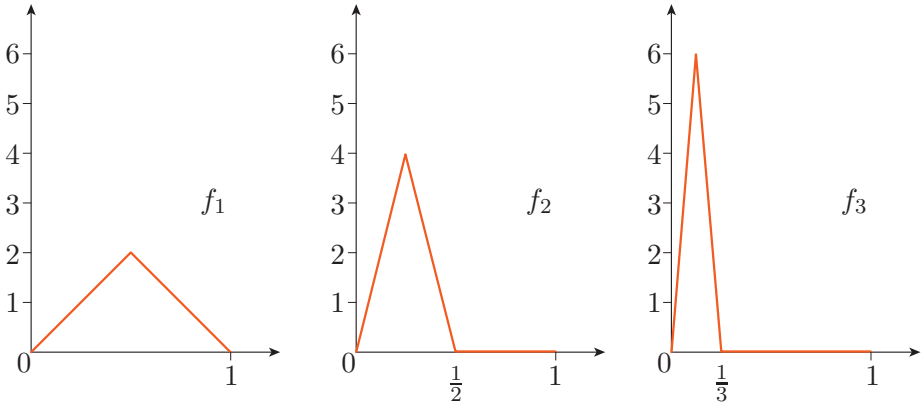
The behaviour at the point $x = 0$ is different: since $f_n(0) = 1$ for each $n \in \mathbb{N}$, $f_n(0) \to 1$ as $n \to \infty$.

So the pointwise limit of the sequence (f_n) is the (discontinuous) function $f : [0, \infty) \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{if } 0 < x < \infty. \end{cases}$$

Solution to Exercise 4.2

(a) The graphs of f_1, f_2 and f_3 are sketched below.



(b) We have $f_n(0) = 0$, for each $n \in \mathbb{N}$, and so $f_n(0) \to 0$ as $n \to \infty$.

If $x > 0$, then we can choose $N \in \mathbb{N}$ so that $\frac{1}{N} < x$. Then, for $n > N$, $f_n(x) = 0$. So $f_n(x) \to 0$ as $n \to \infty$.

Hence the pointwise limit function f of (f_n) is the zero function on $[0, 1]$. Thus $\int_0^1 f(x)\,dx = 0$.

(c) For each $n \in \mathbb{N}$, we find the integral of f_n by noting that the area representing the integral of f_n is a triangle with height $2n$ and width $\frac{1}{n}$. Thus

$$\int_0^1 f_n(x)\,dx = 1,$$

and so

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 1.$$

(d) From part (b), $\int_0^1 \lim_{n\to\infty} f_n(x)\,dx = 0 \neq 1$ and so we conclude that the processes of integration and taking pointwise limits cannot, in general, be interchanged.

Solution to Exercise 4.3

We begin by identifying the pointwise limit of the sequence. For each $x \in [0, 1]$, $1 - (\frac{x}{n})^2 \to 1$ as $n \to \infty$, and so the pointwise limit of the sequence (f_n) is the constant function $f\colon [0, 1] \to \mathbb{R}$ given by $f(x) = 1$ for each $x \in [0, 1]$.

Thus, by Lemma 4.5, our only candidate for the uniform limit of the sequence (f_n) is the function f. To show that it is the uniform limit, we must show that the sequence (M_n), where M_n is given by

$$M_n = \sup\{|f_n(x) - f(x)| : x \in [0, 1]\},$$

is a null sequence.

To prove this, we observe that for each $x \in [0, 1]$ and each $n \in \mathbb{N}$,

$$\begin{aligned} 0 \leq |f_n(x) - f(x)| = |1 - (\tfrac{x}{n})^2 - 1| &= |\tfrac{x}{n}|^2 \\ &\leq \tfrac{1}{n^2}. \end{aligned}$$

Hence for each $n \in \mathbb{N}$,

$$0 \leq M_n = \sup\{|1 - (\tfrac{x}{n})^2 - 1| : x \in [0, 1]\} = \tfrac{1}{n^2}.$$

Since $(\frac{1}{n^2})$ is a basic null sequence, it follows that (M_n) is a null sequence.

Thus $f_n \to f$ uniformly as $n \to \infty$.

Solution to Exercise 4.4

For each $x \in \mathbb{R}$, $\frac{x}{n} \to 0$ as $n \to \infty$, and so the pointwise limit of (f_n) is the zero function given by $f(x) = 0$ for each $x \in \mathbb{R}$. Thus the only possible candidate for the uniform limit of the sequence (f_n) is f.

But $f_n - f$ is unbounded for each n: given any $M > 0$, $(f_n - f)(Mn) = f_n(Mn) - 0 = M$. Thus, for each $n \in \mathbb{N}$,

$$M_n = \sup\{|f_n(x) - f(x)| : x \in \mathbb{R}\} = +\infty$$

and so, in particular, (M_n) is not a real null sequence. Hence (f_n) does not converge uniformly to the zero function on $\mathbb{R}$.

Thus by Lemma 4.5, (f_n) is not uniformly convergent.

Solution to Exercise 4.5

(a) For each $x > 0$, we use the fact that $|\sin y| \leq 1$ for all $y \in \mathbb{R}$ to estimate

$$|f_n(x)| \leq \frac{1}{nx}.$$

Thus, for each $x > 0$, since $(\frac{1}{n})$ is a basic null sequence, we use the Multiple and Squeeze Rules to conclude that $f_n(x) \to 0$ as $n \to \infty$.

In contrast, we have $f_n(0) = 1$ for each $n \in \mathbb{N}$, and so $f_n(0) \to 1$ as $n \to \infty$.

Thus the sequence converges pointwise to the limit function $f\colon [0, \infty) \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{if } x > 0. \end{cases}$$

This function is discontinuous at 0.

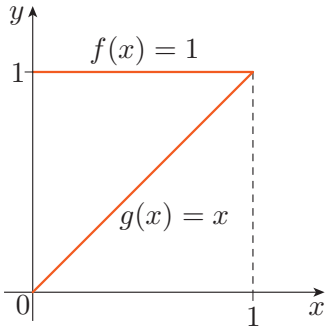
(b) Since

$$\lim_{y \to 0} \frac{\sin y}{y} = 1,$$

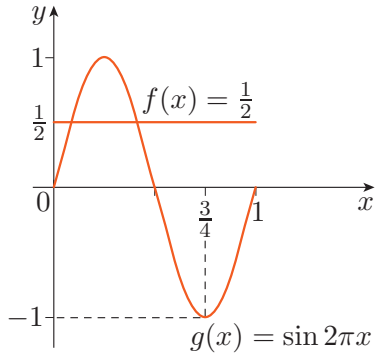
we deduce that for each $n \in \mathbb{N}$ the function f_n is continuous at $x = 0$. From Lemma 4.5 we know that the only possible uniform limit of the sequence (f_n) is the pointwise limit f. But if $f_n \rightrightarrows f$ as $n \to \infty$, then by Theorem 4.7 f would be continuous at 0. We conclude that the sequence (f_n) is not uniformly convergent.

Solution to Exercise 5.1

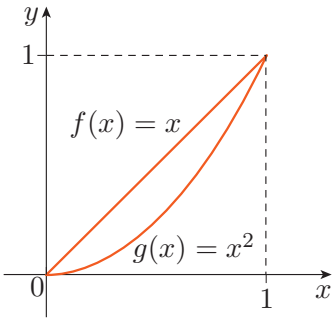
(a) Here the maximum value of $|g(x) - f(x)|$ is 1 (and is attained when $x = 0$); so $d_{\max}(f, g) = 1$.

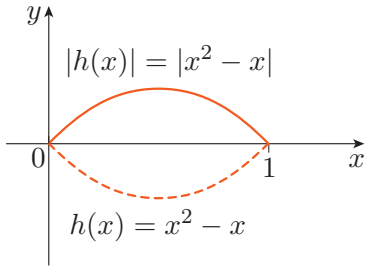


(b) In this case, the furthest $g(x)$ gets from $\frac{1}{2}$ is when $g(x) = -1$ (at $x = \frac{3}{4}$) and so $d_{\max}(f, g) = |(-1) - \frac{1}{2}| = \frac{3}{2}$.



(c) On $[0, 1]$, $g(x) \leq f(x)$ with equality at the endpoints only. Thus the maximum of $|g - f|$ over $[0, 1]$ occurs at the minimum of the function $h(x) = g(x) - f(x) = x^2 - x$ over this domain. This occurs at the point x_0 for which the derivative $h'(x) = 2x - 1$ vanishes, so that $x_0 = \frac{1}{2}$. Thus $d_{\max}(f, g) = |\frac{1}{4} - \frac{1}{2}| = \frac{1}{4}$.





Solution to Exercise 5.2

No, $d_{\min}$ does not define a metric on $C[0, 1]$, because it does not satisfy (M1). For example, if $f(x) = 0$ and $g(x) = x$ for $x \in [0, 1]$, then $d_{\min}(f, g) = 0$ but $f \neq g$.

Solution to Exercise 5.3

Let $h \in C[0,1]$ be given. Then $I_h = I \circ G_h$, where $G_h \colon C[0,1] \to C[0,1]$ is given by

$$G_h(f)(x) = h(x)f(x)$$

and $I \colon C[0,1] \to \mathbb{R}$ is given by

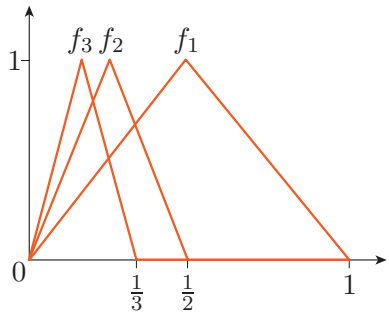
$$I(f) = \int_0^1 f(x)\,dx.$$

In Worked Exercises 5.5 and 5.6, we saw that G_h is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$, and that I is $(d_{\max}, d^{(1)})$-continuous on $C[0,1]$.

Hence, by Proposition 4.4 of Chapter 14, with $X = Y = C[0,1], Z = \mathbb{R}$, $d_X = d_Y = d_{\max}$, $d_Z = d^{(1)}$, $f = G_h$ and $g = I$, we find that I_h is $(d_{\max}, d^{(1)})$-continuous on $C[0,1]$.

Solution to Exercise 5.4

(a) A sketch of the three graphs is shown below.



(b) For each $n \in \mathbb{N}$, $f_n(0) = 0$, and so $f_n(0) \to 0$ as $n \to \infty$.

If $0 < x \leq 1$, then $f_n(x) = 0$ for all $n > \frac{1}{x}$, and hence $f_n(x) \to 0$ as $n \to \infty$.

Thus, for each $x \in [0,1]$, $\lim_{n\to\infty} f_n(x) = 0$.

(c) We use the strategy. Step 1 of the strategy requires us to identify a possible limit. By part (b), the only possible limit for the sequence (f_n) is the pointwise limit function, that is, the zero function $f \colon [0,1] \to \mathbb{R}$ given by $f(x) = 0$ for each $x \in [0,1]$. This is a continuous function, and so we move to Step 3 of the strategy and calculate $d_{\max}(f_n, f)$, for $n \in \mathbb{N}$.

In fact, since $f_n(\frac{1}{2n}) = 1$ and $f_n(x) < 1$ for $x \neq \frac{1}{2n}$, we have, for each $n \in \mathbb{N}$,

$$d_{\max}(f_n, f) = \max\{|f_n(x) - 0| : x \in [0,1]\} = 1.$$

Thus the sequence $(d_{\max}(f_n, f))$ is not a null sequence, and so (f_n) does not converge in $C[0,1]$ for $d_{\max}$.

Chapter 16

Open and closed sets

Introduction

In this chapter we introduce some ideas that help us to describe particular important kinds of subsets that occur in metric spaces. This provides us with some useful terminology that we will make extensive use of in Book F, where we return to our study of metric spaces. Metric spaces can be very abstract objects and having terms for particular types of subsets can help us to understand their structure better.

Our starting point for this is the main tool we have used for investigating and understanding metric spaces, namely convergent sequences. We start this chapter by looking at sets that behave 'nicely' with regard to convergent sequences. In particular, we look at sets that contain the limit points of those convergent sequences lying within them – such sets are known as *closed sets* and they turn out to play an important role in some of the theorems that we will prove in Book F. We investigate the properties of closed sets and look at how to determine them for particular metric spaces.

We also look at another important class of subsets – the *open sets*. These sets occur as the complements of closed sets, and as a result there are close connections between their properties and the properties of closed sets.

Using the notion of *closure*, which we introduce in Section 3, we find a way to associate a 'natural' closed set with any set. It is also possible to associate an open set with any given set and we investigate this in Section 4, where we introduce the notion of the *interior* of a set.

In Section 5 we discuss how to find the *boundary* of an arbitrary set, which consists of those points that are (arbitrarily) near to both the set and its complement.

Section 6 takes a slightly different perspective from the rest of the chapter and looks at two ways of measuring the 'size' of sets. We first look at a metrical notion of large sets – those whose closure equals the whole space. In the second half of the section, we look at a notion of size that is purely set-theoretic and does not use the metric (should one be available). This leads us to discovering that the set of real numbers is strictly 'bigger' than the set of natural numbers.

In the final section, which is not assessed, we discuss how the ideas of this chapter lead naturally to the notion of a *topology*.

Chapter guide

This chapter contains many concepts but, with the exception of the discussion of countable and uncountable, they all rely on the notion of a closed set, which is defined in the first section, and so this is the key idea to understand.

If you are short of time, then you should concentrate your efforts on the first four sections of the chapter; you may wish to omit the last section, which is not directly assessed.

1 Closed sets

In Definition 1.4 of Chapter 15 we defined the notion of the induced or subspace metric for a subset $A \subseteq X$. This is the metric for A given by $d_A(a, b) = d(a, b)$ for $a, b \in A$. From the definition of the subspace metric and the definition of convergence for a sequence in the original metric space (X, d), we can see what it means for a sequence in (A, d_A) to be d_A-convergent:

> A sequence (a_n) in A is d_A-convergent if there is $a \in A$ for which $(d(a_n, a))$ is a real null sequence.

So suppose that (X, d) is a metric space and that A is a subset of X. Let (a_n) be a convergent sequence with respect to the metric d, all of whose terms are in A. Is the limit a of the sequence necessarily in A?

Exercise 1.1

For each of the following metric spaces and sets A, state whether or not the limit of the given convergent sequence is in the set A.

(a) Suppose that the space is $\mathbb{R}$ with the Euclidean metric, the sequence is $(\frac{1}{n})$ and the set is

(i) $A = [0, 1]$

(ii) $A = (0, 1)$.

(b) Suppose that the space is $(\mathbb{R}^2, e_1)$, the plane with the taxicab metric, the sequence $(\mathbf{a}_n)$ is given by $\mathbf{a}_n = (1 + \frac{1}{2n}, 2 - \frac{1}{n})$ for $n \in \mathbb{N}$ and the set is

(i) $A = B_{e_1}[(0, 1), 2]$

(ii) $A = B_{e_1}((0, 1), 2)$.

By Exercise 3.1 of Chapter 14 we know that this sequence converges to $(1, 2)$.

Exercise 1.1 demonstrates that a sequence that is convergent in the original metric space need not be convergent (with respect to the subspace metric) in a subspace that contains all the terms of the sequence.

Subsets of a metric space that (in the subspace metric) do contain the limits of convergent sequences are known as closed sets.

Definition 1.1 *Closed set*

Let (X, d) be a metric space and let $A \subseteq X$.

We say that A is a **d-closed set** if for each d-convergent sequence with terms in A, the limit of the sequence is also in A. That is, whenever there is a d-convergent sequence (a_n) with $a_n \in A$ for each n then $\lim_{n\to\infty} a_n \in A$.

Remarks

1. When it is clear from the context, we may suppress reference to the metric d and just refer to A being **closed**.
2. The definition of closed can be described in terms of the subspace metric on A: the set A is a d-closed set if, and only if, for each sequence (a_n) in A, (a_n) is d_A-convergent if, and only if, (a_n) is d-convergent. (Notice that it is always the case that a d_A-convergent sequence in the subspace is d-convergent in the original space, and so d_A-convergence only has to be tested for those sequences in A that are convergent in the larger space.)

1.1 Examples of closed sets

We will shortly look at closed sets in some specific metric spaces, but first we identify some sets that are always closed.

Lemma 1.2

Let (X, d) be a metric space. The following subsets of X are always d-closed:

1. the empty set, $\emptyset$
2. the whole space, X
3. sets consisting of a single point $\{x\}$ for some $x \in X$.

Proof

1. For the empty set, there are no sequences whose terms lie in it and so there is no condition to test. Hence it is closed. (This is an example of a result that holds purely because there is nothing that can fail to satisfy the condition to be tested.)
2. We need to show that if (a_n) is a sequence in X that is d-convergent (to x, say), then the limit is in X – but this is exactly the definition of convergence in (X, d) and so X is d-closed.
3. Let $x \in X$ be given and let $A = \{x\}$. If (a_n) is a sequence in A, then $a_n = x$ for each $n \in \mathbb{N}$ and so in particular is an eventually constant sequence. Hence by Worked Exercise 3.5 of Chapter 14, (a_n) is d-convergent with limit x, which is in A. Thus A is d-closed as claimed. ■

Our next standard (and slightly more interesting) examples of closed sets are the closed balls in a metric space, which at least shows that our choice of terminology is consistent!

Recall that
$B_d[x, r] = \{y \in X : d(y, x) \leq r\}$.

Proposition 1.3

Let (X, d) be a metric space, let $x \in X$ and let $r \geq 0$. Then the closed ball $B_d[x, r]$ is a d-closed set.

Proof Let (a_n) be a convergent sequence in X with limit a, and suppose that $a_n \in B_d[x, r]$ for each $n \in \mathbb{N}$. In order to show that the ball $B_d[x, r]$ is d-closed we need to show that $a \in B_d[x, r]$. Thus our goal is to show that $d(a, x) \leq r$.

Now, for each $n \in \mathbb{N}$, the Triangle Inequality for d and the fact that $a_n \in B_d[x, r]$ gives

$$d(a, x) = d(x, a) \leq d(x, a_n) + d(a_n, a) \leq r + d(a_n, a).$$

But $(d(a_n, a))$ is a null sequence and so for each $\varepsilon > 0$, there is $n \in \mathbb{N}$ for which $d(a_n, a) < \varepsilon$. Hence

$$d(a, x) < r + \varepsilon, \text{ for each } \varepsilon > 0.$$

In other words, $d(a, x) \leq r$, and so $a \in B_d[x, r]$ and the closed ball $B_d[x, r]$ is d-closed. ■

In particular, closed intervals of the form $[a, b]$ for $a, b \in \mathbb{R}$ with $a \leq b$ are closed sets for the Euclidean metric because
$[a, b] = B_{d^{(1)}}[\frac{1}{2}(a + b), \frac{1}{2}(b - a)]$.

Worked Exercise 1.4

Let $a \in \mathbb{R}$ and let d denote the Euclidean metric for $\mathbb{R}$. Show that $(-\infty, a]$ is a d-closed set.

Solution

Let (x_n) be a d-convergent sequence in $\mathbb{R}$ with limit x, and suppose that $x_n \in (-\infty, a]$ for each $n \in \mathbb{N}$. In order to show that the interval $(-\infty, a]$ is d-closed we need to show that $x \in (-\infty, a]$. Thus our goal is to show that $x \leq a$.

So suppose instead that $x > a$; then $|x - x_n| = x - x_n \geq x - a > 0$ for each $n \in \mathbb{N}$. Thus if we choose $\varepsilon > 0$ for which $x - a > \varepsilon$, then we have $|x - x_n| > \varepsilon$ for each $n \in \mathbb{N}$, contradicting the fact that $x_n \to x$ as $n \to \infty$. Hence $x \leq a$ and $(-\infty, a]$ is d-closed as required.

Exercise 1.2

Let $a \in \mathbb{R}$ and let d denote the Euclidean metric for $\mathbb{R}$. Show that $(-\infty, a)$ is not a d-closed set.

Although the interval $(-\infty, a)$ is not closed for the Euclidean metric, its *complement* is. Before asking you to show this, we recall the definition of the complement of a set.

Definition 1.5 *Complement of a set*

Let A be a subset of a set X. The **complement of A with respect to X** is $A^c = X - A$, the set of elements that are in X but not in A (Figure 1.1).

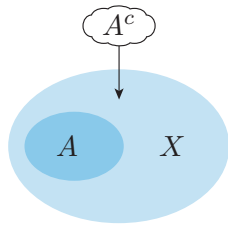


Figure 1.1 A^c is the complement of A in X

Remarks

1. The meaning of A^c depends on the context: we need to know the set X with respect to which we are taking the complement.
2. We can deduce immediately from the definition that $(A^c)^c = A$.

Exercise 1.3

Write down the complement A^c of $A = \{1, 3\}$ with respect to X when

(a) $X = \{1, 2, 3\}$; (b) $X = \{1, 2, 3, 4, 5\}$.

Exercise 1.4

Let $X = \mathbb{R}$ and $A = [0, 1]$. Write the complement A^c of A with respect to X as a union of intervals.

You are now asked to show that the complement of the interval $(-\infty, a)$ is closed for the Euclidean metric.

Exercise 1.5

Let $a \in \mathbb{R}$ and let d denote the Euclidean metric for $\mathbb{R}$. Show that the complement of $(-\infty, a)$ in $\mathbb{R}$ is a d-closed set.

Worked Exercise 1.6

Let (X, d) be a metric space. Show that for every choice of $x \in X$ and $r > 0$, the complement of the 'open' ball $B_d(x, r)$, $X - B_d(x, r)$, is d-closed.

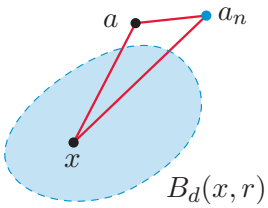


Figure 1.2 Showing that $X - B_d(x,r)$ is closed

Solution

Let $x \in X$ and $r > 0$ and suppose that the sequence (a_n) is such that $a_n \in X - B_d(x,r)$ for each $n \in \mathbb{N}$. That is, for each $n \in \mathbb{N}$, $d(a_n, x) \geq r$. If the sequence (a_n) is convergent for d with limit a, then the Reverse Triangle Inequality (Proposition 2.2 in Chapter 14) gives that for each $n \in \mathbb{N}$,

$$d(x,a) \geq |d(a_n,x) - d(a_n,a)| \geq r - d(a_n,a),$$

as illustrated in Figure 1.2. But $(d(a_n,a))$ is a real null sequence and so

$$d(x,a) \geq r,$$

which means that $a \notin B_d(x,r)$. That is, $a \in X - B_d(x,r)$, as required.

Exercise 1.6

Let (X,d) be a metric space, and $x \in X$ and $r \geq 0$ be given. Show that $A = \{y \in X : d(x,y) \geq r\}$ is a d-closed set.

We have now seen that intervals of the form $[a,b]$, $[a,\infty)$ or $(-\infty,a]$, which we already termed as being 'closed', are indeed closed for the Euclidean metric. Moreover, Exercise 1.5 shows that the complement of what we would usually term an 'open' interval, $(-\infty,a)$, is in fact d-closed: this turns out to be important when we come to define the notion of open sets (for a metric d) later in this chapter.

Worked Exercise 1.7

Show that $\mathbb{Q}$ is not a closed subset of $\mathbb{R}$ for the Euclidean metric.

Solution

To do this it is enough to find a sequence of rational numbers that converges to a point that is not rational. An example of such a sequence is given by (x_n), where x_n is the rational number obtained by taking the first n digits in the usual decimal representation of $\sqrt{2}$.

So $x_1 = 1$, $x_2 = 1.4$, $x_3 = 1.41$, ...

We now look at some other examples of closed sets for some of the metric spaces that we have met in previous chapters.

Closed sets in the plane

We have already seen that, for $\mathbb{R}$ with the usual Euclidean metric, sets of the form $[a, b]$, $(-\infty, a]$ and $[a, \infty)$ (for $a, b \in \mathbb{R}$) are closed. We also know that $\mathbb{R}$ itself is closed, as are sets consisting of a single point of $\mathbb{R}$. We have also seen some examples of sets that are not closed, such as the (infinite) 'open' interval $(-\infty, a)$ (for $a \in \mathbb{R}$), the 'open' interval $(0, 1)$ (see Exercise 1.1) and $\mathbb{Q}$.

We now look at examples of closed sets in the plane $\mathbb{R}^2$ (with the usual notion of distance, $d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$). We already know from Proposition 1.3 that closed balls $B_d[\mathbf{x}, r]$ and the whole of the plane are closed sets for the Euclidean metric, so we are looking for some other examples of closed sets.

Before looking at our first examples, we record a useful lemma concerning convergent real sequences.

Lemma 1.8

Let (a_n) be a convergent sequence of real numbers with limit a (for the usual Euclidean metric). If $m, M \in \mathbb{R}$ are such that for each $n \in \mathbb{N}$, $m \leq a_n \leq M$, then

$$m \leq a \leq M.$$

Exercise 1.7

Prove Lemma 1.8.

Worked Exercise 1.9

Consider the plane $\mathbb{R}^2$ with the usual Euclidean metric d.

(a) Show that

$$\begin{aligned} U &= (0, 1) \times (0, 3) \\ &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : 0 < x_1 < 1 \text{ and } 0 < x_2 < 3\} \end{aligned}$$

is not a d-closed set.

(b) Show that

$$\begin{aligned} H &= [0, 1] \times [0, 3] \\ &= \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1 \text{ and } 0 \leq x_2 \leq 3\} \end{aligned}$$

is a d-closed set.

U and H are illustrated in Figure 1.3.

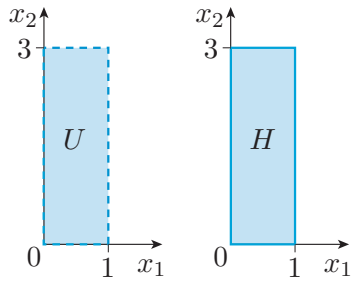


Figure 1.3 The sets U and H

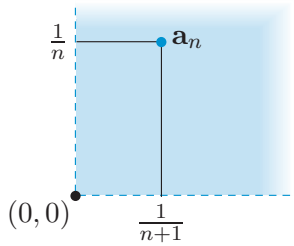


Figure 1.4 $(\mathbf{a}_n) = (\frac{1}{n+1}, \frac{1}{n})$

Solution

(a) We need only find a sequence whose terms lie in U that converges to a point outside U. The points on the 'boundary' of U are not in U and so we try a sequence that converges to a point on the boundary – define the sequence $(\mathbf{a}_n)$ by

$$a_n = (\tfrac{1}{n+1}, \tfrac{1}{n}) \text{ for } n \in \mathbb{N},$$

as shown in Figure 1.4. Clearly, since $0 < \frac{1}{n+1} < 1$ and $0 < \frac{1}{n} \leq 1 < 3$ for each $n \in \mathbb{N}$, $\mathbf{a}_n \in U$ for each $n \in \mathbb{N}$. However, $\mathbf{a}_n \to (0,0) \notin U$ as $n \to \infty$ for the usual Euclidean distance. Hence U is not closed for the Euclidean metric.

(b) In this case, we need to show that whenever $(\mathbf{a}_n)$ is a convergent sequence whose terms all lie in H, then the limit of the sequence (with respect to the Euclidean metric) also lies in H.

So suppose $(\mathbf{a}_n) = ((a_{1,n}, a_{2,n}))$ is such a sequence with limit $\mathbf{a} = (a_1, a_2)$ – we must show that $\mathbf{a} \in H$.

For each $n \in \mathbb{N}$,

$$0 \leq a_{1,n} \leq 1 \text{ and } 0 \leq a_{2,n} \leq 3.$$

Moreover, since $(\mathbf{a}_n)$ is convergent, we know from Theorem 3.15 of Chapter 15 that so are both of the real sequences $(a_{1,n})$ and $(a_{2,n})$ with limits a_1 and a_2, say. We can now immediately use Lemma 1.8 to deduce that

$$0 \leq a_1 \leq 1 \text{ and } 0 \leq a_2 \leq 3,$$

implying that $\mathbf{a} = (a_1, a_2) \in H$ as required.

The results of Worked Exercise 1.9 can be generalised to show that any set of the form $(a,b) \times (c,d)$, often referred to as an *open rectangle*, is not closed for the Euclidean metric on the plane and that any set of the form $[a,b] \times [c,d]$, often referred to as a *closed rectangle*, is closed.

Worked Exercise 1.10

Let $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and $r > 0$. Show that for the plane with the usual Euclidean metric d, the open ball $B_d(\mathbf{x}, r)$ is not closed.

Solution

Let $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and $r > 0$, and consider the sequence $(\mathbf{a}_n)$ given by $\mathbf{a}_n = \left(x_1 + r\left(1 - \frac{1}{n+1}\right), x_2\right)$ for $n \in \mathbb{N}$.

Then for each $n \in \mathbb{N}$,

$$d(\mathbf{a}_n, \mathbf{x}) = \sqrt{r^2(1 - \tfrac{1}{n+1})^2 + 0^2} = r(1 - \tfrac{1}{n+1}) < r$$

and so $\mathbf{a}_n \in B(\mathbf{x}, r)$ for each $n \in \mathbb{N}$.

However,

$$d(\mathbf{a}_n, (x_1 + r, x_2)) \to 0 \text{ as } n \to \infty,$$

meaning that the sequence $(\mathbf{a}_n)$ converges to $(x_1 + r, x_2)$ as $\mathbf{n} \to \infty$ (Figure 1.5). But $d(\mathbf{x}, (x_1 + r, x_2)) = r$ and so $(x_1 + r, x_2) \notin B_d(\mathbf{x}, r)$. Hence $B_d(\mathbf{x}, r)$ is not closed.

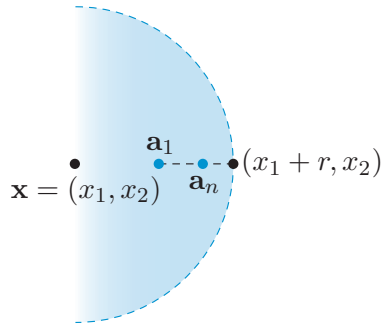


Figure 1.5 Sequence converging to $(x_1 + r, x_2)$

Exercise 1.8

Let $A = \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1\}$.

(a) Sketch A. Be careful to indicate which points lie in A.

(b) Determine whether A is closed for the Euclidean metric, d.

Closed sets for the discrete metric

We now look at closed sets when the discrete metric d_0 is used. We consider an arbitrary subset A of (X, d_0) and test whether d_0-convergent sequences of points from A must converge to a point in A; it turns out that *every subset of X is d_0-closed.*

Recall that for $x, y \in X$, $d_0(x, y) = 1$ unless $x = y$, in which case $d_0(x, y) = 0$.

Worked Exercise 1.11

Let X be a set. Show that every subset of X is d_0-closed.

Solution

Let $A \subseteq X$ be any subset and let (a_n) be a d_0-convergent sequence all of whose terms are in A. We must show that the limit of the sequence, a, say, also is a member of A.

Now, recall that a sequence in X is d_0-convergent if, and only if, it is eventually constant (see Worked Exercise 3.5 and Exercise 3.2 of Chapter 14) and so there is $N \in \mathbb{N}$ such that $a_n = a_N$ for $n \geq N$ and, in particular, a_N is the d_0-limit of the sequence (a_n). But by assumption $a_N \in A$ and so the limit of the sequence lies in A. Hence A is a d_0-closed set as required.

We saw earlier that intervals of the form $(-\infty, a)$, where $a \in \mathbb{R}$, and the set $\mathbb{Q}$ of rational numbers are not closed in the metric space $(\mathbb{R}, d^{(1)})$. However, it follows from the result of Worked Exercise 1.11 that each of these sets *is* closed in the space $(\mathbb{R}, d_0)$. Thus a set can be closed with respect to one metric but not with respect to another.

Closed sets in Cantor space

Recall from Chapter 15 that the Cantor space $\mathbf{C}$ consists of all infinite sequences of zeros and ones, and that we define the Cantor metric $d_{\mathbf{C}}$ by

$$d_{\mathbf{C}}(\mathbf{x},\mathbf{y}) = \begin{cases} 0, & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-n}, & \text{if } \mathbf{x} \text{ and } \mathbf{y} \text{ first differ at their } n\text{th term.} \end{cases}$$

In view of Proposition 1.3, we know that closed balls in Cantor space are $d_{\mathbf{C}}$-closed sets. But combining the results of Worked Exercise 2.4 and Exercise 2.3 of Chapter 15, we see that for $\mathbf{a} \in \mathbf{C}$ and $n \in \mathbb{N}$,

$$B_{d_{\mathbf{C}}}(\mathbf{a}, 2^{-n}) = B_{d_{\mathbf{C}}}[\mathbf{a}, 2^{-(n+1)}] = \{(a_1, a_2, \ldots, a_n, c_1, c_2, \ldots) : (c_i) \in \mathbf{C}\}.$$

Thus open balls in $\mathbf{C}$ are also $d_{\mathbf{C}}$-closed sets! This turns out to be a symptom of the fact that Cantor space is not connected – a topic that we take up in Book F, Chapter 21.

Closed sets in $(C[0,1], d_{\max})$

We give a couple of examples of closed subsets in $C[0,1]$.

Worked Exercise 1.12

Let $M \geq 0$ be given and let

$$A_M = \{f \in C[0,1] : |f(x)| \leq M \text{ for each } x \in [0,1]\},$$

so A_M consists of those continuous functions on $[0,1]$ whose values lie in the range $[-M, M]$.

Show that A_M is a closed subset of $(C[0,1], d_{\max})$.

Solution

Suppose that (f_n) is a $d_{\max}$-convergent sequence of functions in $C[0,1]$ with $f_n \in A_M$ for each n. We must show that the limit of the sequence, f, say, also lies in A_M. That is, for each $x \in [0,1]$, we must show $|f(x)| \leq M$. So suppose $x \in [0,1]$; then for each $n \in \mathbb{N}$ we can use the Triangle Inequality to find an upper bound for $|f(x)|$ involving f_n as follows:

$$|f(x)| = |f(x) - f_n(x) + f_n(x)| \leq |f(x) - f_n(x)| + |f_n(x)|.$$

But $|f(x) - f_n(x)| \leq d_{\max}(f_n, f)$ and $f_n \in A_M$, and so $|f_n(x)| \leq M$. Hence

$$|f(x)| \leq d_{\max}(f_n, f) + M \text{ for each } n \in \mathbb{N}.$$

But $(d_{\max}(f_n, f))$ is a null sequence and the value of $|f(x)|$ does not depend on n, and so we conclude that

$$|f(x)| \leq M.$$

Since x was an arbitrary real number in $[0,1]$, this implies that $f \in A_M$. We conclude that A_M is $d_{\max}$-closed.

An alternative way to show this would be to observe that $A_M = B_{d_{\max}}[g, M]$ where $g \in C[0,1]$ is the 'zero function' given by $g\colon [0,1] \to \mathbb{R}$, $g(x) = 0$, and so A_M is $d_{\max}$-closed by Proposition 1.3. We will see yet another (more flexible) way of obtaining this result in the next subsection.

Exercise 1.9

Let $M \in \mathbb{R}$ be given and let

$$B_M = \{f \in C[0,1] : f(0) = M\},$$

so B_M consists of those continuous functions on $[0,1]$ that take the value M at zero.

Show that B_M is a closed subset of $(C[0,1], d_{\max})$.

Exercise 1.10

Let $M \geq 0$ and $a \in [0,1]$ be given, and let

$$C = \{f \in C[0,1] : |f(a)| \leq M\},$$

so C consists of those continuous functions on $[0,1]$ that take a value in $[-M, M]$ at the point a.

Show that C is a closed subset of $(C[0,1], d_{\max})$.

1.2 The algebra of closed sets

We now have some examples of closed sets in several different kinds of metric space but, as you may have already concluded, it can be a bit tedious to verify directly from the definition of closed sets whether a given set is closed. In this subsection we look at how closed sets can be combined; in the same way that we use combination rules to combine known convergent sequences in $(\mathbb{R}, d^{(1)})$, this gives us a way to combine known closed sets to form new closed sets. In particular we investigate when unions and intersections of closed sets are also closed.

Before starting we record a useful result concerning convergent sequences in metric spaces.

Lemma 1.13

Let (X, d) be a metric space and $(a_n)_{n=1}^{\infty}$ be a sequence in X. If (a_n) is d-convergent, then any subsequence of (a_n) is also d-convergent and has the same limit as the original sequence.

Proof Let (a_n) be d-convergent with limit $a \in X$ and let $(a_{n_i})_{i=1}^{\infty}$ be a subsequence of (a_n). We must show that $(d(a_{n_i}, a))_{i=1}^{\infty}$ is a (real) null sequence.

Consider the real sequence (d_n) where $d_n = d(a_n, a)$; this is a real null sequence and the sequence $(d(a_{n_i}, a))_{i=1}^{\infty} = (d_{n_i})_{i=1}^{\infty}$ is a subsequence of it. Hence by Proposition 1.25 of Chapter 13, (d_{n_i}) is also convergent with limit 0. Thus $(d(a_{n_i}, a))_{i=1}^{\infty}$ is a (real) null sequence. ■

This result tells us that if, for a given sequence, we can find two subsequences that converge to different limits, then the original sequence must be divergent.

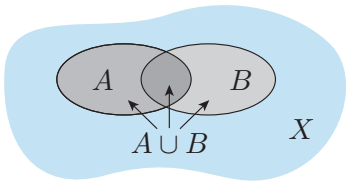


Figure 1.6 The union of A and B

We can now prove our first result concerning combining closed sets. It concerns the union of two sets, as shown in Figure 1.6. Recall that the union of two sets A and B, denoted $A \cup B$, is given by $A \cup B = \{x \in X:$ either $x \in A$ or $x \in B\}$. The 'or' in the definition of union is inclusive in that elements of the union of A and B can appear in both A and B.

Proposition 1.14

Let (X, d) be a metric space and let A and B be d-closed subsets of X. Then $A \cup B$ is also d-closed.

Proof Let (x_n) be a d-convergent sequence all of whose terms lie in $A \cup B$, and suppose it has limit x: we must show that $x \in A \cup B$. Then for at least one of A and B there is a subsequence of (x_n), $(x_{n_i})_{i=1}^{\infty}$, all of whose terms lie in it: suppose it is A. (So $x_{n_i} \in A$ for each $i \in \mathbb{N}$.) Then by Lemma 1.13, since (x_n) is convergent, so is the subsequence (x_{n_i}) and it has the same limit x. But $x_{n_i} \in A$ for each i and A is closed and so $x \in A$, which in particular means that $x \in A \cup B$ as required. ■

This result immediately tells us that the union of a finite number of closed sets is also closed. For a finite union of sets $A_1 \cup A_2 \cup \cdots \cup A_n$ can be written as the union of two sets $(A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \cup A_n$ and we can use induction to deduce that if all the sets $A_1, \ldots, A_n$ are closed, then so is their union.

Thus, for example, a finite union of closed intervals is itself closed (with respect to the usual Euclidean metric).

Recall that a closed interval $[a, b] = B_d[\frac{1}{2}(a+b), \frac{1}{2}(b-a)]$ for the Euclidean metric d.

Worked Exercise 1.15

Let $a, b \in \mathbb{R}$ with $a \leq b$. Show that the complement of the open interval (a, b) is closed for the usual Euclidean metric.

Solution

The complement of the open interval (a, b) is

$$\mathbb{R} - (a, b) = (-\infty, a] \cup [b, \infty).$$

We saw in Worked Exercise 1.4 and Exercise 1.5 that the intervals $(-\infty, a]$ and $[b, \infty)$ are closed for the Euclidean metric on $\mathbb{R}$. Hence Proposition 1.14 implies that their union is also closed for the Euclidean metric.

Exercise 1.11

Let (X, d) be a metric space and let $x_1, x_2, \ldots, x_n \in X$. Show that the finite set $\{x_1, x_2, \ldots, x_n\}$ is d-closed.

Exercise 1.12

Let d denote the usual Euclidean metric on $\mathbb{R}$. Give examples of sets $A_1, A_2, A_3, \ldots$ in $\mathbb{R}$ such that A_n is d-closed for each n but $\cup_{n=1}^{\infty} A_n = \{a \in X : a \in A_n \text{ for some } n \in \mathbb{N}\}$ is not d-closed.

Hint: you may find it helpful to consider sets that each consist of a single term of an appropriately chosen convergent sequence.

Exercise 1.13

Let $K_0 = [0, 1]$ and suppose that for each $n \in \mathbb{N}$, K_n is obtained from the set K_{n-1} by replacing each interval I in K_{n-1} with two closed intervals of one-third the length of I, one starting at the left endpoint of I and the other ending at the right endpoint of I.

(a) Write down K_1 and K_2 as a union of intervals.

(b) Prove that K_n is closed for the Euclidean metric on $\mathbb{R}$ for each $n \in \mathbb{N}$.

The next result concerns the intersection of an arbitrary collection of closed sets, so before giving it, let us recall the definition of intersection. We already know what the intersection of two sets $A \cap B$ means – it consists of those points that are members of both A and B – but we are going to need to look at intersections of many more than two sets, so we should agree the terminology.

Definition 1.16 *Intersection*

Let $\{A_i : i \in I\}$ be a collection of subsets of a given set X.

The **intersection** of the collection of sets, written $\bigcap_{i\in I} A_i$ or $\cap\{A_i : i \in I\}$, is defined by

$$\bigcap_{i\in I} A_i = \{x \in X : x \in A_i \text{ for each } i \in I\}.$$

Thus the intersection consists of exactly those points that the sets A_i have in common.

The set I is called an **index set** for the collection.

Remarks

1. The index set can be finite or infinite and gives a way of labelling the sets so that we can refer to them directly.
2. For each $i \in I$, $\bigcap_{i\in I} A_i \subseteq A_i$. That is, the intersection of a collection of sets is always contained inside each of the sets in the collection.

Proposition 1.17

Let (X, d) be a metric space and let $\{A_i : i \in I\}$ be a collection of d-closed sets. Then $\bigcap_{i\in I} A_i$ is d-closed.

Proof Let (x_n) be a d-convergent sequence all of whose terms lie in $\bigcap_{i\in I} A_i$, and suppose that the limit of the sequence is x – we must show that for each $i \in I$, $x \in A_i$.

So fix $i \in I$. Since for each $n \in \mathbb{N}$, $x_n \in \bigcap_{i\in I} A_i$, it is certainly the case that $x_n \in A_i$ for each $n \in \mathbb{N}$.

But A_i is d-closed and (x_n) is a d-convergent sequence lying within A_i. Hence the limit of the sequence (x_n), x, also lies in A_i, as required. ■

We have shown that closed balls $B_d[x, r]$ are closed sets in a metric space (and that open balls, $B_d(x, r)$, need not be), but what about spheres?

Theorem 1.18

Let (X, d) be a metric space, let $a \in X$ and let $r \geq 0$. Then the sphere $S_d(a, r) = \{x \in X : d(a, x) = r\}$ is a d-closed set.

Exercise 1.14

Prove Theorem 1.18.

Example 1.19 *The middle-third Cantor set*

We now take a first look at an example of a closed set in the unit interval (with the usual Euclidean distance) that plays an important role in the final chapter of the module, Chapter 24.

As in Exercise 1.13, let $K_0 = [0, 1]$ and form K_1 from K_0 by removing the open middle-third of K_0 – the open interval $(\frac{1}{3}, \frac{2}{3})$, meaning that $K_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, the union of two (closed) intervals of length $\frac{1}{3}$. We continue in this way, constructing a set K_n for each $n \in \mathbb{N}$. Thus K_n is formed by removing the open middle-third of each interval in K_{n-1} and so K_n consists of 2^n closed intervals each with length $\frac{1}{3^n}$. Note that the sets K_n are the same as those in Exercise 1.13 and so are closed for the Euclidean metric.

We define the (middle-third) Cantor set K_C by $K_C = \bigcap_{n=1}^{\infty} K_n$. Then K_C is non-empty, since it contains the points 0 and 1, and, by Proposition 1.17, it is closed. However, notice that it does not contain any intervals because if x and y are in K_C with $x \neq y$ then, on choosing n so that $3^{-n} < |x - y|$, we see that x and y cannot lie in the same interval from K_n and so there is a point z lying between x and y that is not in K_C (Figure 1.7). Hence, in particular, if (x, y) is an (open) interval then there is $x < z < y$ with $z \notin K_C$. Thus K_C contains no intervals – it is an example of what is called a *totally disconnected set*. We investigate this and related ideas more fully in Chapter 21.

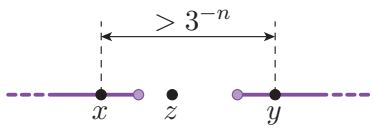


Figure 1.7 The point z lies between x and y but is not in K_C

Worked Exercise 1.20

Fix $M \geq 0$ and let

$$D = \{f \in C[0, 1] : |f(x)| \leq M \text{ for each } x \in [0, \tfrac{1}{2}]\}.$$

Show that D is $d_{\max}$-closed.

Solution

The result of Exercise 1.10 tells us that for each $a \in [0, \frac{1}{2}]$, the set C_a given by

$$C_a = \{f \in C[0, 1] : |f(a)| \leq M\}$$

is $d_{\max}$-closed. Now

$$D = \bigcap_{a \in [0, \frac{1}{2}]} C_a$$

and so is the intersection of a collection of $d_{\max}$-closed sets. Hence by Proposition 1.17, D is $d_{\max}$-closed.

Exercise 1.15

For $M \in \mathbb{N}$, let

$$F_M = \left\{ f \in C[0,1] : f\left(\tfrac{1}{M}\right) = 0 \right\}.$$

(a) Show that F_M is a closed set in $(C[0,1], d_{\max})$.

(b) Give an example of a function f lying in $\bigcap_{M\in\mathbb{N}} F_M$.

(c) Show that $\bigcap_{M\in\mathbb{N}} F_M$ is a closed set in $(C[0,1], d_{\max})$.

2 Open sets

In the last section we extended the idea of a closed interval in $\mathbb{R}$ to the notion of a closed set in a metric space. We also know about open intervals in $\mathbb{R}$ such as (a,b) for $a < b$, and in this section we see how we can define a more general notion of being 'open' in a metric space.

Our starting point is based on some observations made in Worked Exercise 1.15 and Exercise 1.5 – namely that the 'open' intervals (a,b) and $(-\infty, a)$ both have complements that are closed sets in $\mathbb{R}$ with the usual Euclidean metric. This leads to an indirect way of defining the notion of 'open' for a general metric space in terms of it being the complement of a closed set.

Definition 2.1 *Open set*

Let (X,d) be a metric space and let $A \subseteq X$. Then A is said to be an **open set for the metric** d if the complement of A in X, $X - A$, is a d-closed set.

Remarks

1. When we wish to emphasise the particular metric d on X, we say that A is a d-**open set** or that the set A is **open** in (X,d).
2. Since the complements of intervals of the form (a,b) and $(-\infty,a)$ are closed for the usual Euclidean metric on $\mathbb{R}$, open intervals in $\mathbb{R}$ are open for the Euclidean metric. (Intervals of the form (a,∞) also have a closed complement.)

As we did with closed sets, we start by identifying some sets that are always open in a metric space – the proof of the following lemma makes use of Lemma 1.2.

Lemma 2.2

Let (X, d) be a metric space. The following subsets of X are d-open:

1. the empty set, $\emptyset$
2. the whole space, X.

Proof

1. The complement of the empty set in X is $X - \emptyset$, which equals X. From Lemma 1.2, we know that X is d-closed and so $\emptyset$ is d-open.
2. The complement of X in X is $X - X$, which equals the empty set $\emptyset$. From Lemma 1.2, we know that $\emptyset$ is d-closed and so X is d-open. ■

Thus the sets $\emptyset$ and X are both open and closed in any metric space (X, d). Later in this chapter we will see further examples of sets that are both open and closed in a particular metric space, $(\mathbf{C}, d_{\mathbf{C}})$.

We now look at open balls, $B_d(x, r)$, in a metric space and confirm that they are, indeed, d-open sets!

Proposition 2.3

Let (X, d) be a metric space, let $a \in X$ and let $r > 0$. Then the (open) ball $B_d(a, r) = \{x \in X : d(x, a) < r\}$ is a d-open set.

Proof In order to show that the ball $B_d(a, r)$ is open, we must show that $X - B_d(a, r)$ is d-closed. But, by the result of Worked Exercise 1.6, $X - B_d(a, r)$ is d-closed. Hence $B_d(a, r)$ is d-open as required. ■

2.1 Examples of open sets

We can take the complements of the examples of closed sets in various metric spaces that we found in Subsection 1.1 to obtain examples of open sets.

Example 2.4

Consider the plane $\mathbb{R}^2$ with the usual Euclidean metric d. Then the result of Worked Exercise 1.9 tells us that the complement of $H = [0, 1] \times [0, 3]$ is a d-open set. That is,

$$\begin{aligned}\mathbb{R}^2 - H &= \mathbb{R}^2 - \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1 \text{ and } 0 \leq x_2 \leq 3\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 < 0 \text{ or } x_1 > 1 \text{ or } x_2 < 0 \text{ or } x_2 > 3\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 < 0\} \cup \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 1\} \\ &\qquad \cup \{(x_1, x_2) \in \mathbb{R}^2 : x_2 < 0\} \cup \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 3\}\end{aligned}$$

is open for the Euclidean metric on the plane.

Similarly, Exercise 1.8 tells us that the complement of $A = \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1\}$ is an open set for the Euclidean metric on the plane. Thus

$$\begin{aligned}\mathbb{R}^2 - A &= \mathbb{R}^2 - \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_1 \leq 1\} \\ &= \{(x_1, x_2) \in \mathbb{R}^2 : x_1 < 0\} \cup \{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 1\}\end{aligned}$$

is open.

Example 2.5

The situation for the discrete metric is interesting. We saw in Worked Exercise 1.11 that if a set X is given the discrete metric, d_0, then every subset of X is d_0-closed. This in turn means that every subset of X is also d_0-open: for if $A \subseteq X$, then $X - A$ is d_0-closed and so A must be open. Thus the discrete metric on a set X gives an example of a metric space where every subset is both open and closed!

Example 2.6

We saw several examples of closed sets for $(C[0,1], d_{\max})$ in Subsection 1.1 and, by taking complements, we immediately obtain examples of open sets. For example, in Exercise 1.9, we saw that B_M given by $B_M = \{f \in C[0,1] : f(0) = M\}$ for some $M \in \mathbb{R}$ is a $d_{\max}$-closed set. Hence for each $M \in \mathbb{R}$, $C[0,1] - B_M = \{f \in C[0,1] : f(0) \neq M\}$ is a $d_{\max}$-open set.

Exercise 2.1

Let $a \in \mathbb{R}$. Show that the interval (a, ∞) is open in $\mathbb{R}$ for the usual Euclidean metric and explain why $[a, \infty)$ is not.

Exercise 2.2

Let $a < b$. Explain why the sets $(a, b]$ and $[a, b]$ are not open for the Euclidean metric.

Exercise 2.2 highlights an important point: sets in a metric space need not be open or closed – the interval $(a, b]$ is neither open nor closed for the usual Euclidean metric.

Let us consider briefly the metric space $(C[0,1], d_{\max})$. Open sets in this space are more difficult to visualise than those in Euclidean spaces.

Exercise 2.3

Let $M \geq 0$ be given and let

$$C_M = \{f \in C[0,1] : \text{there is } x \in [0,1] \text{ for which } |f(x)| > M\},$$

so C_M consists of those continuous functions on $[0,1]$ that take a value outside the range $[-M, M]$ for some point in $[0,1]$.

Show that C_M is an open subset of $(C[0,1], d_{\max})$.

Working directly from the definition of open sets requires us to look at the complement of the set in question and show that this is closed for the metric. This is a rather indirect approach and it would be helpful if we had a way of showing a set is open without having to resort to looking at the complement each time. A clue towards doing this is provided by the following observation concerning open balls in a metric space.

Theorem 2.7

Let (X, d) be a metric space, let $a \in X$ and $r > 0$, and consider the open ball $B_d(a, r)$. Then, whenever $x \in B_d(a, r)$, we can find an $s > 0$ such that

$$B_d(x, s) \subseteq B_d(a, r).$$

We prove this shortly.

Example 2.8

Consider the open ball $B_{d^{(1)}}(2, 1)$ in $(\mathbb{R}, d^{(1)})$, that is, the open interval $(1, 3)$. Given any $x \in (1, 3)$, we can find $B_{d^{(1)}}(x, s)$; that is, the open interval $(x - s, x + s)$ such that

$$(x - s, x + s) \subseteq (1, 3).$$

For instance, if $x = 2.99$, then we may choose $s = 0.005$, since $(2.985, 2.995) \subseteq (1, 3)$.

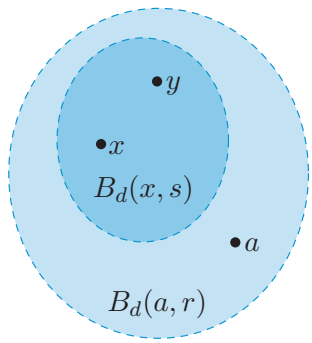


Figure 2.1
$B_d(x, s) \subseteq B_d(a, r)$

The properties of a metric space were first listed in Chapter 14, Definition 2.1.

We now prove Theorem 2.7.

Proof of Theorem 2.7 Let $x \in B_d(a, r)$, so $d(a, x) < r$. Now let $s = r - d(a, x) > 0$. Then, for $y \in B_d(x, s)$, we may use property (M3) (the Triangle Inequality) for d to conclude that

$$d(a, y) \leq d(a, x) + d(x, y) < d(a, x) + s = r.$$

That is, $y \in B_d(a, r)$. Thus $B_d(x, s) \subseteq B_d(a, r)$, as shown in Figure 2.1. ■

The closer to the boundary of the ball $B_d(a, r)$ that x lies, the smaller the radius s may be: the crucial point is that *there is always room enough*.

This observation about open balls extends to give us a useful characterisation of open sets.

Theorem 2.9

Let (X, d) be a metric space and let A be a subset of X. Then A is d-open if, and only if, for each $a \in A$ there is $r > 0$ so that $B_d(a, r) \subseteq A$.

Proof *If:*

We first suppose that for each $a \in A$, there is $r > 0$ so that $B_d(a, r) \subseteq A$. Our goal is to show that the complement of A, $X - A$, is d-closed.

So let (x_n) be a d-convergent sequence with limit x for which $x_n \in X - A$ for each n. If we show that $x \in X - A$, then $X - A$ is d-closed. So suppose instead that $x \notin X - A$, that is, $x \in A$.

Then by our hypothesis there is $s > 0$ so that $B_d(x, s) \subseteq A$. Since by Proposition 2.3 $B_d(x, s)$ is d-open, its complement, C, say, is d-closed. Now $(X - A) \subseteq C$ and so $x_n \in C$ for each $n \in \mathbb{N}$. But since C is d-closed and each of the terms $x_n \in C$, it follows that $x \in C$. But this is impossible because $C \cap B_d(x, s) = \emptyset$. Thus our assumption that $x \notin X - A$ is false and so $x \in X - A$. This shows that $X - A$ is a d-closed set and so, by definition, A is a d-open set.

Only if:

We now suppose that A is an open set and show that for each $a \in A$ there is $r > 0$ so that $B_d(a, r) \subseteq A$.

Let $B = X - A$ be the complement of A. Then since A is d-open, B is a d-closed set. If the claimed conclusion does not hold, then there is $a \in A$ such that for each $r > 0$, $B_d(a, r) \cap B \neq \emptyset$. In particular, for this choice of $a \in A$ and each $n \in \mathbb{N}$ we can choose $a_n \in B_d(a, \frac{1}{n}) \cap B$.

Thus, in particular, $a_n \in B$ for each $n \in \mathbb{N}$ and $d(a, a_n) \leq \frac{1}{n}$. This means that $(d(a, a_n))$ is a null sequence and hence that (a_n) is d-convergent with limit a.

But B is d-closed and (a_n) is a d-convergent sequence and so it must be the case that $a \in B$, which is a contradiction.

Thus no such a can exist and A has the claimed property. ■

Worked Exercise 2.10

Show that

$$S = (-1, 1) \times (-1, 1)$$

is open in $\mathbb{R}^2$ for the Euclidean metric.

Solution

There are two possible ways to approach this: either show that the complement of S is a closed set or use Theorem 2.9. We use Theorem 2.9 and show that for each $\mathbf{a} \in S$, we can find an $r > 0$ so that $B_d(\mathbf{a}, r) \subseteq S$ (where d denotes the Euclidean metric for the plane).

So let $\mathbf{a} = (a_1, a_2) \in S$. Then both $|a_1| < 1$ and $|a_2| < 1$.

Let $r = \min\{1 - |a_1|, 1 - |a_2|\}$, so that r is the minimum distance that the point $\mathbf{a}$ is from the 'boundary' of S (see Figure 2.2).

We claim that $B_d(\mathbf{a}, r) \subseteq S$. To show this, suppose $\mathbf{x} = (x_1, x_2) \in B_d(\mathbf{a}, r)$ – we must show that $\mathbf{x} \in S$. Now by the Triangle Inequality (for modulus), the fact that $|x_1 - a_1| \leq d(\mathbf{x}, \mathbf{a}) < r$ and the definition of r, we can find an upper bound for $|x_1|$:

$$\begin{aligned} |x_1| &= |x_1 - a_1 + a_1| \leq |x_1 - a_1| + |a_1| \\ &< r + |a_1| \leq (1 - |a_1|) + |a_1| = 1. \end{aligned}$$

Similarly, we can find an upper bound for $|x_2|$:

$$\begin{aligned} |x_2| &= |x_2 - a_2 + a_2| \leq |x_2 - a_2| + |a_2| \\ &< r + |a_2| \leq (1 - |a_2|) + |a_2| = 1. \end{aligned}$$

Thus both $|x_1| < 1$ and $|x_2| < 1$ and so $\mathbf{x} \in S$. We conclude that S is an open set.

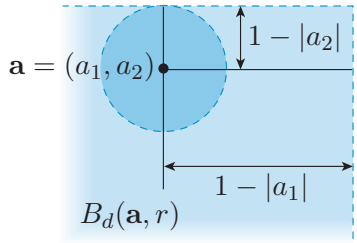


Figure 2.2 Minimum distance of $\mathbf{a}$ from the boundary of S

Worked Exercise 2.11

Let c be a point in $[0, 1]$, and consider the set A_c of all functions $f \in C[0, 1]$ whose value $f(c)$ at c lies in the real open interval $(-1, 1)$. If we think of $(-1, 1)$ as a vertical 'gate' sited at $x = c$, then f belongs to A_c only if it passes through the gate:

$$A_c = \{f \in C[0, 1] : |f(c)| < 1\}.$$

Figure 2.3 shows a function $g \in A_c$ and a function $h \notin A_c$.

Show that A_c is an open set in $(C[0, 1], d_{\max})$.

Solution

We show that for any $g \in A_c$ there is $r > 0$ so that $B_{d_{\max}}(g, r) \subseteq A_c$, which by Theorem 2.9 implies that A_c is $d_{\max}$-open.

Since $g(c) \in (-1, 1)$, we can find $r > 0$ with $r < 1 - |g(c)|$. Then

$$(g(c) - r, g(c) + r) \subseteq (-1, 1).$$

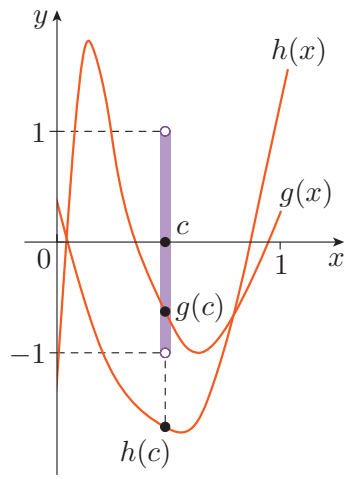


Figure 2.3 The function g lies in the set A_c of Worked Exercise 2.11 whereas the function h does not

With this value of r, consider the open ball $B_{d_{\max}}(g, r)$. If $h \in B_{d_{\max}}(g, r)$, then $|h(x) - g(x)| < r$ for each $x \in [0, 1]$, so that in particular $|h(c) - g(c)| < r$. Thus

$$-1 < g(c) - r < h(c) < g(c) + r < 1,$$

which shows that $h \in A_c$. Thus if $h \in B_{d_{\max}}(g, r)$, then $h \in A_c$ and so $B_{d_{\max}}(g, r) \subseteq A_c$.

Since g is an arbitrary point in A_c, it follows from Theorem 2.9 that A_c is an open subset of $(C[0, 1], d_{\max})$.

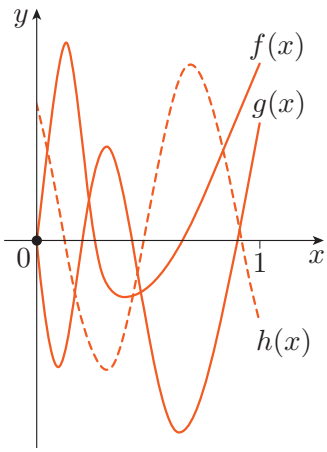


Figure 2.4 The functions f and g lie in S, but the function h does not

It is not too difficult to find examples of sets in $(C[0, 1], d_{\max})$ that are not open. For example, consider the set S of all functions whose graphs pass through the origin:

$$S = \{f \in C[0, 1] : f(0) = 0\}.$$

Figure 2.4 shows two functions f and g in S, and a function h not in S.

Exercise 2.4

Show that the set S is not open.

We saw when looking at examples of closed sets in Subsection 1.1 that in Cantor space $(\mathbf{C}, d_{\mathbf{C}})$ the open balls $B_{d_{\mathbf{C}}}(\mathbf{a}, r)$ are $d_{\mathbf{C}}$-closed sets. There is a symmetric result concerning closed balls in Cantor space.

Exercise 2.5

Let $(\mathbf{C}, d_{\mathbf{C}})$ be the Cantor space with the Cantor metric. Show that for $\mathbf{a} \in \mathbf{C}$ and $n \in \mathbb{N}$, the closed ball $B_{d_{\mathbf{C}}}[\mathbf{a}, 2^{-n}]$ is a $d_{\mathbf{C}}$-open set.

In Chapter 14 we observed that a sequence (a_n) in X converges to $a \in X$ if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $a_n \in B_d(a, \varepsilon)$ whenever $n > N$. The next exercise asks you to interpret this result in a slightly different way.

Exercise 2.6

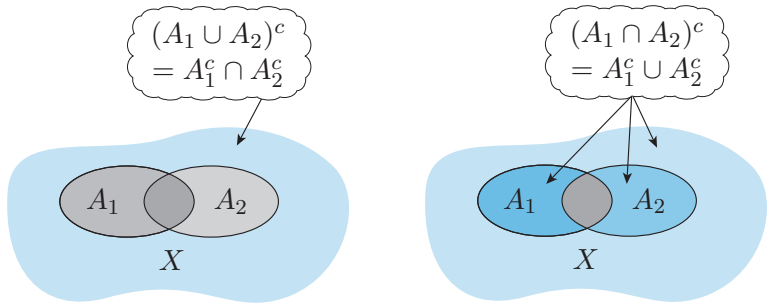
Let (a_n) be a sequence in the metric space (X, d) that does not converge to the point $a \in X$. Show that there is an open ball $B_d(a, \varepsilon)$ (for some $\varepsilon > 0$) such that there is a subsequence (a_{m_n}) with $a_{m_n} \notin B_d(a, \varepsilon)$ for each $n \in \mathbb{N}$.

2.2 The algebra of open sets

In the same way as we investigated how closed sets could be combined through union and intersection in Subsection 1.2, we now look at how open sets can be combined. Given that open sets are defined in terms of the complements of closed sets, one would expect that the rules for their combination can be derived from those for closed sets – and indeed, this is the case. To do this, we make use of *De Morgan's Laws*, which give identities useful for calculating complements of unions and intersections of sets. In the simplest case, when we have just two sets A_1 and A_2 lying within a 'universal set' X, these laws take the following form:

$$X - (A_1 \cup A_2) = (A_1 \cup A_2)^c = A_1^c \cap A_2^c = (X - A_1) \cap (X - A_2);$$
$$X - (A_1 \cap A_2) = (A_1 \cap A_2)^c = A_1^c \cup A_2^c = (X - A_1) \cup (X - A_2).$$

These are illustrated in Figure 2.5.



(a) (b)

Figure 2.5 (a) $X - (A_1 \cup A_2)$; (b) $X - (A_1 \cap A_2)$

Augustus De Morgan (1806–1871) contributed to set theory and mathematical logic, and was a distinguished populariser of mathematics. He was a founding member of the London Mathematical Society, which today, despite its name, is a leading UK-wide learned society for mathematics.

Exercise 2.7

For $X = \{1, 2, \ldots, 12\}$, let $A_1 = \{x \in X : x \text{ is even}\}$ and let $A_2 = \{x \in X : x \text{ is divisible by } 3\}$. Verify De Morgan's Laws for the sets A_1 and A_2.

The most general form of De Morgan's Laws applies to arbitrary families of subsets of a set X.

Theorem 2.12 *De Morgan's Laws*

Let $\{A_i : i \in I\}$ be a family of subsets of X. Then:

First Law

$$X - \left(\bigcup_{i\in I} A_i\right) = \left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c = \bigcap_{i\in I}(X - A_i);$$

Second Law

$$X - \left(\bigcap_{i\in I} A_i\right) = \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c = \bigcup_{i\in I}(X - A_i).$$

Although it is not difficult, we do not prove this result.

De Morgan's First Law can be remembered as 'the complement of the union is (identical to) the intersection of the complements' and De Morgan's Second Law as 'the complement of the intersection is the union of the complements'.

Exercise 2.8

Write out De Morgan's Laws when the index set is $I = \{1, 2, 3\}$.

We now use De Morgan's Laws together with Propositions 1.14 and 1.17 to investigate how open sets are combined.

Theorem 2.13

In a metric space, the intersection of any two open sets is an open set.

Proof Let (X, d) be a metric space, and suppose that U and V are d-open subsets of X. Then, by definition, $X - U$ and $X - V$ are d-closed sets.

Consider $U \cap V$: to show that this is a d-open set, it is enough to show that $(U \cap V)^c = X - (U \cap V)$ is d-closed.

However, by De Morgan's Second Law,

$$X - (U \cap V) = (X - U) \cup (X - V)$$

and both $X - U$ and $X - V$ are d-closed. Hence, by Proposition 1.14, so is their union, implying that $U \cap V$ is d-open as required. ■

Remarks

1. If U_1, U_2, U_3 are open in the metric space (X, d), then

   $$U_1 \cap U_2 \cap U_3 = (U_1 \cap U_2) \cap U_3$$

   and we may use Theorem 2.13 first to deduce that $U_1 \cap U_2$ is open and then to deduce that $(U_1 \cap U_2) \cap U_3$ is open. Thus, the intersection of any three open sets is open.

   This method extends to any *finite* intersection of open sets and allows us to conclude that if $U_1, U_2, \ldots, U_n$ are open sets in a metric space (X, d), then so is

   $$U = U_1 \cap U_2 \cap \cdots \cap U_n.$$

2. However, this result does *not* extend to the intersection of *infinitely many* open sets in a metric space. For example, consider the following collection of open intervals in $\mathbb{R}$ with its usual metric $d(x, y) = |x - y|$:

   $$A_1 = (-1, 1), A_2 = (-\tfrac{1}{2}, \tfrac{1}{2}), A_3 = (-\tfrac{1}{3}, \tfrac{1}{3}), \ldots, A_n = (-\tfrac{1}{n}, \tfrac{1}{n}), \ldots,$$

   illustrated in Figure 2.6. Each set A_n (being an open interval) is open, but

   $$\cap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap A_3 \cap \cdots$$

   is the set of real numbers lying in *all* the intervals $(-\frac{1}{n}, \frac{1}{n})$ for all $n \in \mathbb{N}$, which is the singleton set $\{0\}$. But $\{0\}$ is not an open set, since any d-open ball about 0 contains points other than 0.

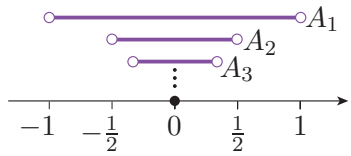


Figure 2.6 $A_n = (-\frac{1}{n}, \frac{1}{n})$

Since Proposition 1.17 tells us that arbitrary intersections of closed sets are closed and De Morgan's Laws have the effect of turning complements of intersections into unions of complements, we would expect that the union of *any* collection of open sets is an open set, irrespective of whether the collection is finite or infinite.

We first remind ourselves of the meaning of the union of a collection of sets.

Definition 2.14 *Union*

Let $\{A_i : i \in I\}$ be a collection of subsets of a given set X.

The **union** of the collection of sets, written $\bigcup_{i \in I} A_i$ or $\cup\{A_i : i \in I\}$, is defined by

$$\bigcup_{i \in I} A_i = \{x \in X : \text{there is } i \in I \text{ for which } x \in A_i\}.$$

Thus the union consists of exactly those points that appear in at least one of the sets A_i.

We can now state our result concerning unions of open sets.

Theorem 2.15

In a metric space, the union of any collection of open sets is open.

Exercise 2.9

Prove Theorem 2.15.

Any *finite* non-empty set of real numbers $A = \{a_1, a_2, \ldots, a_n\}$ is not open in $(\mathbb{R}, d^{(1)})$: in any open ball centred on a point in A, there are points not in A. Similarly, if (a_n) is a strictly increasing sequence of real numbers,

$$a_1 < a_2 < a_3 < \cdots,$$

then the set

$$A = \{a_n : n \in \mathbb{N}\}$$

cannot be open in $(\mathbb{R}, d^{(1)})$. To see this, consider any point a_k of A: for any $r > 0$, $B_{d^{(1)}}(a_k, r)$ contains points not in A. Thus, for example, the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers is not open in $(\mathbb{R}, d^{(1)})$.

Worked Exercise 2.16

Show that the set $\mathbb{Q}$ of rational numbers is not open in $(\mathbb{R}, d^{(1)})$.

Solution

Let $x \in \mathbb{Q}$. We show that for each $r > 0$, there is a point $y \in B_{d^{(1)}}(x, r)$ such that $y \notin \mathbb{Q}$ – that is, each open ball centred on a rational number must contain an irrational number.

So, let $r > 0$ and consider $y = x + \frac{\sqrt{2}}{k}$, where $k \in \mathbb{N}$ is chosen so that $\frac{\sqrt{2}}{k} < r$. It follows that $y \in B_{d^{(1)}}(x, r)$. We now need to verify that y is irrational. Suppose $y = x + \frac{\sqrt{2}}{k}$ is rational. Then, on rearranging, we obtain

$$k(y - x) = \sqrt{2}.$$

This implies that $\sqrt{2}$ is rational, which is false, so our assumption must be false. Thus, y is irrational and so is not in $\mathbb{Q}$.

Hence $B_{d^{(1)}}(x, r) \not\subseteq \mathbb{Q}$ and it follows, since r was arbitrary, that $\mathbb{Q}$ is not open.

One consequence of Theorem 2.15 is that if we can write a given set as a union of open sets, then we immediately know that it is open and we do not have to find open balls around each point to verify this. For example, let

$$J = \{(x, 0) \in \mathbb{R}^2 : -2 \leq x \leq 2\},$$

and let

$$U = \{\mathbf{u} \in \mathbb{R}^2 : d^{(2)}(\mathbf{u}, \mathbf{j}) < 1 \text{ for some } \mathbf{j} \in J\}.$$

Thus U is a rectangle with 'rounded ends' (Figure 2.7), and it may be intuitively clear to you that U is open. We could construct a proof of this from the definition of an open set, but there is no need: because U has been defined as the union of the infinite collection of open balls $B_{d^{(2)}}((x, 0), 1)$ where $-2 \leq x \leq 2$, we can deduce immediately from Theorem 2.15 that U is open.

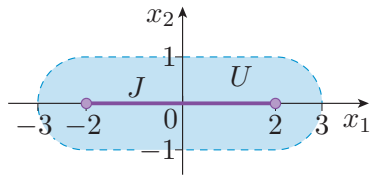


Figure 2.7 Representation of U

This principle is also effective for more abstract metric spaces, such as $(C[0, 1], d_{\max})$. For example, we saw in Worked Exercise 2.11 that for each $c \in [0, 1]$, the set $A_c = \{f \in C[0, 1] : |f(c)| < 1\}$ is $d_{\max}$-open. Hence we can deduce immediately from Theorem 2.15 that the set

$$A = \{f \in C[0, 1] : \text{there is a } c \in [0, 1] \text{ for which } |f(c)| < 1\}$$

is a $d_{\max}$-open set, since A is the union of the infinite collection of sets A_c for $c \in [0, 1]$.

Exercise 2.10

Show that the following are open sets in the metric space $(\mathbb{R}^2, d)$ when d is the usual Euclidean metric for the plane.

(a) $U = \{(x, y) \in \mathbb{R}^2 : y > 1\}$

(b) $V = \{(x, y) \in \mathbb{R}^2 : (x - 1)^2 + (y - 1)^2 < 1 \text{ and } (x - 2)^2 + (y - 2)^2 < 1\}$

(c) The set W of points $(x, y) \in \mathbb{R}^2$ such that there is a point $(a, b) \in \mathbb{R}^2$ with $a^2 + b^2 = 1$ for which

$$(x - a)^2 + (y - b)^2 < \tfrac{1}{4}.$$

We finish this section by summarising various ways of showing that a given set is open or closed or neither.

Strategy: to determine whether sets are open or closed

Let (X, d) be a metric space and let $A \subseteq X$. We list the various techniques that you have seen for showing whether A is open, closed or neither. Remember that A can be both open and closed or could be neither.

- To show A is closed:
  - show that the limits of every convergent sequence with terms all lying in A lie in A; or
  - show that A is the union of a finite number of known closed sets; or
  - show that A is the intersection of an arbitrary collection of known closed sets; or
  - show that the complement of A is open.
- To show A is open:
  - show that the complement of A (in X) is closed; or
  - show that for each $a \in A$, there is $r > 0$ so that $B_d(a, r) \subseteq A$; or
  - show that A is the intersection of a finite number of known open sets; or
  - show that A is the union of an arbitrary collection of known open sets.
- To show A is not closed:
  - show that there is a convergent sequence (a_n) with $a_n \in A$ for each n and $\lim_{n\to\infty} x_n \notin A$; or
  - show that the complement of A is not open.
- To show A is not open:
  - show that there is a point $a \in A$ with no open ball $B_d(a, \varepsilon) \subseteq A$ – that is, show that for each open ball $B_d(a, \varepsilon)$, we have $B_d(a, \varepsilon) \cap (X - A) \neq \emptyset$; or
  - show that the complement of A is not closed.

3 Closure

Given a metric space (X, d) we have identified two particular important types of sets lying in X: the closed sets, defined via sequences, and the open sets, which are sets that are complements of closed sets. However, many subsets of metric spaces are neither open nor closed and it would be helpful to have a natural way of associating an open and a closed set with any given subset of a metric space – this is what we investigate in this section.

We have seen that an interval such as $(0, 1]$ is an example of a subset of $\mathbb{R}$ that is neither open nor closed for the Euclidean metric; we now use it to investigate how we could associate a natural open and closed set with it.

First, let us consider how to associate a reasonable closed set with $(0, 1]$. With luck, one such set may already have sprung to your mind, namely the interval $[0, 1]$. This is certainly a closed set for the Euclidean metric, and it is not so very different from our starting set A in that it only contains a single extra point, 0. Similarly, if we try to associate a natural open set with the set A, then the open interval $(0, 1)$ springs to mind.

Exercise 3.1

For each of the following subsets of $\mathbb{R}$, state the open and closed sets that you would naturally associate with them (for the usual Euclidean metric):

(a) $[0, 1)$; (b) $[0, 1]$; (c) $(0, 1)$; (d) $\{0\}$.

In the preceding exercise, we saw that a natural associated closed set always contained the starting set and was the 'smallest' closed set containing the starting set. Similarly we saw that a natural associated open set is the 'largest' open set that lies inside the starting set. This is not a coincidence. Suppose that we have been given a set A in a metric space (X, d) and we wish to find a closed set that contains it. There is always at least one such closed set, namely the space X itself. Now suppose that we have a family of closed sets $\{E_i : i \in I\}$, each of which contains A. Then:

1. $A \subseteq \bigcap_{i\in I} E_i$, since $A \subseteq E_i$ for each $i \in I$.
2. Proposition 1.17 tells us that $\bigcap_{i\in I} E_i$ is also a closed set.

So, given any collection of closed sets containing A, their intersection is also a closed set that contains A.

Consider now the collection

$$\{E \subseteq X : A \subseteq E \text{ and } E \text{ is } d\text{-closed in } X\}.$$

Then, in view of the preceding, the set F that is defined by

$$F = \bigcap \{E \subseteq X : A \subseteq E \text{ and } E \text{ is } d\text{-closed in } X\}$$

is a closed set that contains A. Moreover if E is a closed set that contains A, then E occurs in the intersection defining F and so $A \subseteq F \subseteq E$. In

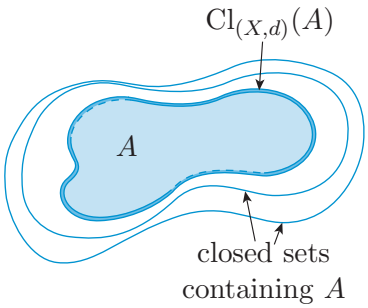


Figure 3.1 The closure of A

other words, F is the smallest closed set that contains A. This is our sought-after definition of the *closure* of a set A.

Definition 3.1 *Closure of a set*

Let (X, d) be a metric space and let $A \subseteq X$.

We define the **closure of A in (X, d)**, $\mathrm{Cl}_{(X,d)}(A)$, to be the smallest d-closed set in X that contains A (Figure 3.1). That is,

$$\mathrm{Cl}_{(X,d)}(A) = \bigcap \{E \subseteq X : A \subseteq E \text{ and } E \text{ is } d\text{-closed in } X\}.$$

Remarks

1. When it is clear from the context we usually just write $\mathrm{Cl}(A)$ and suppress mention of the metric space (X, d).
2. The closure of a set is closed.
3. It is always the case that $A \subseteq \mathrm{Cl}_{(X,d)}(A)$.
4. If F is a d-closed set and $A \subseteq F$, then $\mathrm{Cl}_{(X,d)}(A) \subseteq F$.

We can go through a similar process to find a suitable definition for the 'largest open set' inside a given set and we do this in Section 4.

The next couple of exercises give you the opportunity to explore closure.

Exercise 3.2

Show that for every metric space (X, d),

$$\mathrm{Cl}_{(X,d)}(\emptyset) = \emptyset \text{ and } \mathrm{Cl}_{(X,d)}(X) = X.$$

Exercise 3.3

Determine $\mathrm{Cl}_{(\mathbb{R},d^{(1)})}((0, 1))$.

3.1 Elementary properties of closures

The following simple properties of closures, which follow immediately from the definition, are very useful.

Theorem 3.2

Let (X, d) be a metric space and let $A, B \subseteq X$ with $A \subseteq B$. Then

$$\mathrm{Cl}_{(X,d)}(A) \subseteq \mathrm{Cl}_{(X,d)}(B).$$

Proof Suppose that $A \subseteq B$. Then certainly $B \subseteq \mathrm{Cl}(B)$ and so $\mathrm{Cl}(B)$ is a closed set that contains A. Hence

$$A \subseteq \mathrm{Cl}(A) = \bigcap \{E \subseteq X : A \subseteq E \text{ and } E \text{ is } d\text{-closed in } X\} \subseteq \mathrm{Cl}(B),$$

as required. ■

We now prove that closed sets are precisely those sets that are equal to their closure.

Theorem 3.3

Let (X, d) be a metric space and let $A \subseteq X$. Then A is closed if, and only if, $\mathrm{Cl}_{(X,d)}(A) = A$.

Proof We already know that, by its definition, $\mathrm{Cl}(A)$ is a closed set. So certainly if $\mathrm{Cl}(A) = A$, then A is also a closed set.

Now suppose that A is closed. We must show that $\mathrm{Cl}(A) = A$. We know that it is always the case that $A \subseteq \mathrm{Cl}(A)$, and so it is enough to show that $\mathrm{Cl}(A) \subseteq A$. But if A is a closed set, then it is a closed set that contains A and so

$$A \subseteq \bigcap \{E \subseteq X : A \subseteq E \text{ and } E \text{ is closed}\} \subseteq A.$$

Thus we conclude that $\mathrm{Cl}(A) = A$. ■

Remarks

1. A consequence of this theorem is that if A is not a closed set, then A is a *proper* subset of $\mathrm{Cl}(A)$: $\mathrm{Cl}(A) - A \neq \emptyset$.
2. Since $\mathrm{Cl}(A)$ is a closed set,
$$\mathrm{Cl}(\mathrm{Cl}(A)) = \mathrm{Cl}(A).$$

Worked Exercise 3.4

Find the closure of

$$A = \{f \in C[0,1] : f(\tfrac{1}{n}) = 0 \text{ for each } n \in \mathbb{N}\}$$

in the metric space $(C[0,1], d_{\max})$.

Solution

We saw in part (c) of Exercise 1.15 that A is closed for $d_{\max}$ and so, by Theorem 3.3, $\mathrm{Cl}(A) = A$.

Exercise 3.4

Let $I \subseteq [0,1]$ be a non-empty set. Find the closure of

$$A = \{f \in C[0,1] : f(x) = 0 \text{ for } x \in I\}$$

in the metric space $(C[0,1], d_{\max})$.

3.2 Further properties of closures

See Theorem 3.2.

Suppose that we have two subsets of a metric space. We have already observed that if $A \subseteq B$ then $\mathrm{Cl}(A) \subseteq \mathrm{Cl}(B)$. But suppose we do not know whether one set is contained in the other. What is the relationship between the closure of their union and the union of their closures? What happens when we look at the intersection of their closures? The next theorem provides some of the answers.

Theorem 3.5

Let (X, d) be a metric space and let $A, B \subseteq X$. Then:

(a) $\mathrm{Cl}_{(X,d)}(A \cup B) = \mathrm{Cl}_{(X,d)}(A) \cup \mathrm{Cl}_{(X,d)}(B)$

(b) $\mathrm{Cl}_{(X,d)}(A \cap B) \subseteq \mathrm{Cl}_{(X,d)}(A) \cap \mathrm{Cl}_{(X,d)}(B)$.

Proof We first prove part (a) of the theorem.

We start by showing that

$$\mathrm{Cl}(A \cup B) \subseteq \mathrm{Cl}(A) \cup \mathrm{Cl}(B).$$

We know that $A \subseteq \mathrm{Cl}(A)$ and $B \subseteq \mathrm{Cl}(B)$ and so

$$A \cup B \subseteq \mathrm{Cl}(A) \cup \mathrm{Cl}(B).$$

In particular, since $\mathrm{Cl}(A) \cup \mathrm{Cl}(B)$ is a closed set (being the union of two closed sets) that contains $A \cup B$, it follows that

$$\mathrm{Cl}(A \cup B) \subseteq \mathrm{Cl}(A) \cup \mathrm{Cl}(B).$$

It remains only to show that

$$\mathrm{Cl}(A) \cup \mathrm{Cl}(B) \subseteq \mathrm{Cl}(A \cup B).$$

To do this, notice that $A \subseteq A \cup B \subseteq \mathrm{Cl}(A \cup B)$ and similarly $B \subseteq A \cup B \subseteq \mathrm{Cl}(A \cup B)$. Hence

$$\mathrm{Cl}(A) \subseteq \mathrm{Cl}(\mathrm{Cl}(A \cup B)) = \mathrm{Cl}(A \cup B), \text{ by Theorem 3.3,}$$

and

$$\mathrm{Cl}(B) \subseteq \mathrm{Cl}(\mathrm{Cl}(A \cup B)) = \mathrm{Cl}(A \cup B).$$

Therefore

$$\mathrm{Cl}(A) \cup \mathrm{Cl}(B) \subseteq \mathrm{Cl}(A \cup B).$$

Thus

$$\mathrm{Cl}(A) \cup \mathrm{Cl}(B) = \mathrm{Cl}(A \cup B),$$

as required.

You are asked to verify part (b) in Exercise 3.5. ■

Exercise 3.5

Let (X, d) be a metric space and let $A, B \subseteq X$. Prove that

$$\mathrm{Cl}_{(X,d)}(A \cap B) \subseteq \mathrm{Cl}_{(X,d)}(A) \cap \mathrm{Cl}_{(X,d)}(B).$$

You are asked to show that the set inclusion in part (b) can be strict in Exercise 3.6.

3.3 An alternative way of characterising closure

In this subsection we look at an alternative way of characterising the closure of a set – one that is more 'hands-on', in that it involves identifying points in the space that have a particular property. This alternative characterisation gives us another way of understanding the closure of a set.

Definition 3.6 *Closure point*

Let A be a subset of the metric space (X, d).

A point $x \in X$ is a **d-closure point** of A in X if for each $r > 0$, $B_d(x, r)$ meets the set A; that is, $B_d(x, r) \cap A \neq \emptyset$.

The idea is that a closure point is a point that (from the point of view of the metric) is 'very close' to A in that if x is a closure point, then there is a sequence of points in A that converges to it and so for each $r > 0$, the set A meets $B_d(x, r)$: $A \cap B_d(x, r) \neq \emptyset$. In particular, if $a \in A$, then a is certainly a closure point of A (regardless of the choice of metric), as shown in Figure 3.2.

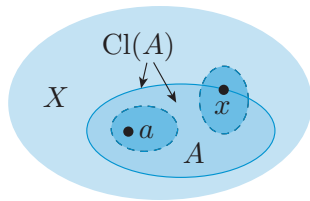


Figure 3.2 Closure points of A

Worked Exercise 3.7

Let (X, d) be a metric space and $A \subseteq X$. Show that $x \in X$ is a d-closure point of A in X if, and only if, there is a sequence of points from A, (a_n), say, that is d-convergent to x.

Solution

We need to show that:

1. If there is a sequence of points from A, (a_n), that is d-convergent to $x \in X$, then x is a d-closure point of A in X.
2. If $x \in X$ is a d-closure point of A in X, then there is a sequence of points from A, (a_n), say, that is d-convergent to x.

So, first, suppose that there is a sequence of points from A, (a_n), that is d-convergent to $x \in X$ and let $r > 0$ be given – we must show that $B_d(x, r) \cap A \neq \emptyset$. However, we know that $(d(a_n, x))$ is a real null sequence and so, in particular, there is $N \in \mathbb{N}$ for which $d(a_n, x) < r$ whenever $n > N$. Hence $a_n \in B_d(x, r)$ for $n > N$ and so $B_d(x, r) \cap A \neq \emptyset$, as required.

Conversely, suppose that $x \in X$ is a d-closure point of A in X; then for each $r > 0$, $B_d(x, r) \cap A \neq \emptyset$. In particular, for each $n \in \mathbb{N}$, we can find $a_n \in B_d(x, \frac{1}{n}) \cap A$. Consider the sequence (a_n): for each $n \in \mathbb{N}$, $a_n \in A$ and $d(a_n, x) < \frac{1}{n}$ and so $(d(a_n, x))$ is a real null sequence. Hence (a_n) is a sequence in A that is d-convergent to x, as required.

The next result tells us that the set of closure points of a given set is the same as the closure of that set, which justifies our choice of terminology.

Theorem 3.8

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Cl}_{(X,d)}(A) = \{x \in X : x \text{ is a } d\text{-closure point of } A \text{ in } X\}.$$

Proof Let $F = \{x \in X : x \text{ is a closure point of } A\}$ – we must show that $F = \mathrm{Cl}(A)$.

The proof proceeds in three steps:

1. Show that $A \subseteq F$.
2. Show that F is closed and so $\mathrm{Cl}(A) \subseteq F$.
3. Show that $F \subseteq \mathrm{Cl}(A)$.

Hence we start by recalling that each point $a \in A$ is a closure point of A and so certainly $A \subseteq F$.

To show that F is closed: suppose that (x_n) is a d-convergent sequence of points in F that converges to $x \in X$, say. We show that $x \in F$.

Since each point $x_n \in F$ is a closure point of A, then for each $n \in \mathbb{N}$, there is $a_n \in A$ for which $d(a_n, x_n) < \frac{1}{n}$. Hence by the Triangle Inequality, for each $n \in \mathbb{N}$,

$$d(a_n, x) \leq d(a_n, x_n) + d(x_n, x) < \tfrac{1}{n} + d(x_n, x).$$

But $(\frac{1}{n})$ and $(d(x_n, x))$ are both real null sequences and so $(d(a_n, x))$ is a real null sequence. In particular, for each $r > 0$, there is $n \in \mathbb{N}$ for which $d(a_n, x) < r$ and so $a_n \in B_d(x, r) \cap A \neq \emptyset$. Thus x is a closure point of A, that is, $x \in F$ and F is closed as claimed.

It only remains to show that $F \subseteq \mathrm{Cl}(A)$. To do this it is enough to show that for each $x \in F$, if E is a closed set that contains A, then $x \in E$. But this is straightforward: since x is a closure point of A, we can find a sequence (a_n) of elements of A that converges to x. But $A \subseteq E$ and so (a_n) is a sequence of elements of E that converges to x. Hence, since E is closed, $x \in E$ as required. ■

Worked Exercise 3.9

Show that for the Euclidean metric on $\mathbb{R}$, $\mathrm{Cl}(\mathbb{Q}) = \mathrm{Cl}(\mathbb{Q}^c) = \mathbb{R}$.

Solution

We begin by showing that $\mathrm{Cl}(\mathbb{Q}) = \mathbb{R}$. We know that $\mathbb{Q} \subseteq \mathrm{Cl}(\mathbb{Q})$. So, since $\mathbb{R} = \mathbb{Q} \cup \mathbb{Q}^c$, we need show only that the irrationals are closure points of $\mathbb{Q}$ to conclude, via Theorem 3.8, that $\mathrm{Cl}(\mathbb{Q}) = \mathbb{R}$.

Let $x \in \mathbb{Q}^c$ and suppose that $r > 0$; then $B(x, r) = (x - r, x + r)$. By the density property of the reals, between any two distinct real numbers there is a rational number and so $(x - r, x + r)$ meets $\mathbb{Q}$. Thus $(x - r, x + r) \cap \mathbb{Q} \neq \emptyset$ and so x is a closure point of $\mathbb{Q}$. Hence $\mathbb{Q}^c \subseteq \mathrm{Cl}(\mathbb{Q})$, as required.

See the *Handbook*.

Similar arguments can be used to show that $\mathrm{Cl}(\mathbb{Q}^c) = \mathbb{R}$.

Exercise 3.6

By considering $\mathbb{Q}$ and $\mathbb{Q}^c$ for the Euclidean metric on $\mathbb{R}$, show that $\mathrm{Cl}(A \cap B)$ may be a *proper* subset of $\mathrm{Cl}(A) \cap \mathrm{Cl}(B)$.

Worked Exercise 3.10

Let $A = \{\frac{1}{n} : n \in \mathbb{N}\} = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$. Find $\mathrm{Cl}(A)$ for the Euclidean metric on $\mathbb{R}$.

Solution

Since $A \subseteq \mathrm{Cl}(A)$, it remains only to decide which points of A^c belong to $\mathrm{Cl}(A)$. Now $A \subseteq [0, 1]$, a closed set, and so, by Theorems 3.2 and 3.3,

$$\mathrm{Cl}(A) \subseteq \mathrm{Cl}([0, 1]) = [0, 1].$$

Hence we need only decide which points in $A^c \cap [0,1]$ are closure points of A.

Now $A^c \cap [0,1] = \{0\} \cup \bigcup_{n=1}^{\infty} (\frac{1}{n+1}, \frac{1}{n})$ and so if $x \in A^c \cap [0,1]$, then either $x = 0$ or there is $n \in \mathbb{N}$ for which $\frac{1}{n+1} < x < \frac{1}{n}$.

Suppose that $x \in A^c \cap [0,1]$.

If $x = 0$, then for each $r > 0$ there is $n \in \mathbb{N}$ for which $\frac{1}{n} < r$, and so $B(0,r) \cap A = (-r, r) \cap A \neq \emptyset$. Hence any ball centred on 0 intersects A. Thus 0 is a closure point of A.

If $x \in (\frac{1}{n+1}, \frac{1}{n})$ for some $n \in \mathbb{N}$, then for $0 < r \leq \min\{x - \frac{1}{n+1}, \frac{1}{n} - x\}$, we have $(x-r, x+r) \cap A = \emptyset$ (Figure 3.3). Hence, $B(x,r) = (x-r, x+r)$ does not intersect A for such an r, and so x is not a closure point of A.

Therefore $\mathrm{Cl}(A) = A \cup \{0\}$.

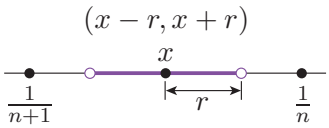


Figure 3.3 $x \in (\frac{1}{n+1}, \frac{1}{n})$

In Worked Exercise 1.11 we showed that for the discrete metric d_0, every subset of a space X is d_0-closed. In particular,

$$\mathrm{Cl}_{(\mathbb{R},d_0)}(\mathbb{Q}) = \mathbb{Q} \neq \mathbb{R} = \mathrm{Cl}_{(\mathbb{R},d^{(1)})}(\mathbb{Q})$$

and so different metrics can give rise to different closures.

Exercise 3.7

Let (a_n) be a strictly increasing sequence of real numbers that is bounded above. If $A = \{a_n : n \in \mathbb{N}\}$, find $\mathrm{Cl}(A)$ for the Euclidean metric on $\mathbb{R}$.

Recall that (a_n) is strictly increasing if $a_n < a_{n+1}$ for all $n \in \mathbb{N}$.

Hint: recall that an increasing sequence of real numbers that is bounded above converges to a limit $a \in \mathbb{R}$, with $a \geq a_n$ for each n.

Exercise 3.8

Let $X = \{(x_1, x_2) : x_1, x_2 \in \mathbb{Q}\}$. Show that for the usual Euclidean metric on the plane, $\mathrm{Cl}(X) = \mathbb{R}^2$.

We can relate closure to sequences.

Corollary 3.11

Let (X,d) be a metric space and let A be a subset of X. A point a belongs to the closure $\mathrm{Cl}_{(X,d)}(A)$ of A if, and only if, there is a sequence (a_n) in A that converges to a.

Proof Theorem 3.8 tells us that the closure of A coincides with the set of closure points of A, and the result of Worked Exercise 3.7 tells us that a point a is a closure point if, and only if, there is a sequence (a_n) in A that converges to a. ■

3.4 Closed sets and subspaces

The ability to associate a closed set with an arbitrary set by taking the closure gives us a way of identifying the closed subsets of a metric subspace (A, d_A) in terms of the closed sets of the original metric space (X, d).

Recall that d_A denotes the subspace (or induced) metric for A.

Theorem 3.12

Let (X, d) be a metric space and let $A \subseteq X$.

A set $E \subseteq A$ is d_A-closed if, and only if, there is a set F in X that is d-closed and for which

$$E = A \cap F.$$

In other words, closed sets in subspaces are just the intersections of the subspace with closed sets from the original space.

Proof Let F be d-closed in X. We show that $A \cap F$ is d_A-closed – that is, the limit of each d_A-convergent sequence in $A \cap F$ is in $A \cap F$.

Suppose that (a_n) is a d_A-convergent sequence in A for which $a_n \in A \cap F$ for each n. Then the limit of the sequence (for d_A), a, say, is in A and $(d_A(a_n, a))$ is a real null sequence. But $d_A(a_n, a) = d(a_n, a)$ for each $n \in \mathbb{N}$ and so a_n d-converges to a in X. But F is closed and so $a \in F$. Hence $a \in A \cap F$ and we conclude that $A \cap F$ is d_A-closed.

Now suppose that E is d_A-closed: we show that there is F in X that is d-closed for which $E = A \cap F$. Let $F = \mathrm{Cl}_{(X,d)}(E)$; this is certainly a d-closed set and we claim that $E = A \cap F$. We know that $E \subseteq A \cap \mathrm{Cl}_{(X,d)}(E)$ and so if we show that $A \cap \mathrm{Cl}_{(X,d)}(E) \subseteq E$, then we have finished.

So suppose that we can find $a \in (A \cap \mathrm{Cl}_{(X,d)}(E)) - E$; that is, suppose that $A \cap \mathrm{Cl}_{(X,d)}(E) \not\subseteq E$. Then $a \in A - E$ and $a \in \mathrm{Cl}_{(X,d)}(E)$. In particular, by Corollary 3.11 we can find a sequence (a_n) in E that d-converges to a. But, since $E \subseteq A$, it follows that each point $a_n \in A$ and so $d_A(a_n, a)$ is defined and equals $d(a_n, a)$, implying that $(d_A(a_n, a))$ is a real null sequence. Hence (a_n) is a sequence that d_A-converges to a. Since E is d_A-closed, this means that $a \in E$, which is a contradiction. We conclude that $A \cap \mathrm{Cl}_{(X,d)}(E) \subseteq E$. ■

4 Interior

In this section we return to the problem of associating an appropriate open set with a given subset of a metric space.

Recall that in Exercise 3.1 we looked at ways of associating an open set with various given subsets of $\mathbb{R}$ for the Euclidean metric, and the conclusions were as follows:

1. with $[0,1)$, $[0,1]$ and $(0,1)$, we associated the open set $(0,1)$
2. with $\{0\}$, we associated the (open) set $\emptyset$.

In all four cases, we found that the natural associated open set is the 'largest' open set that lies inside the starting set. This is not a coincidence. Suppose that we have been given a set A in a metric space (X,d) and we wish to find an open set inside it. There is always at least one such open set, namely the empty set, $\emptyset$. Now suppose that we have a family of open sets $\{U_i : i \in I\}$ each of which is contained in A; then

1. $\bigcup_{i\in I} U_i \subseteq A$, since $U_i \subseteq A$ for each $i \in I$
2. Theorem 2.15 tells us that $\bigcup_{i\in I} U_i$ is also an open set.

So, given any collection of open sets contained in A, their union is also an open set contained in A.

Consider now the collection

$$\{U : U \subseteq A \text{ and } U \text{ is } d\text{-open in } X\}.$$

Then, in view of the preceding, the set V that is defined by

$$V = \bigcup\{U : U \subseteq A \text{ and } U \text{ is } d\text{-open in } X\}$$

is an open set that is contained in A. Moreover if W is an open set that is contained in A, then W occurs in the union defining V and so $W \subseteq V \subseteq A$. In other words, V is the largest open set that is contained in A. This is our sought-after definition of what is known as the *interior* of a set A.

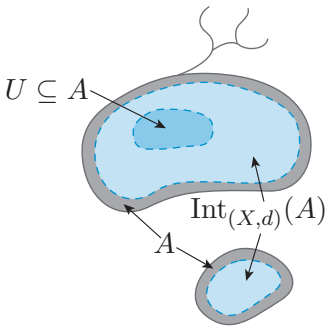


Figure 4.1 The interior of A

Definition 4.1 *Interior of a set*

Let (X,d) be a metric space and suppose $A \subseteq X$. We define the **interior of A in (X,d)**, $\mathrm{Int}_{(X,d)}(A)$, to be the largest d-open set in X that is contained in A (Figure 4.1). That is,

$$\mathrm{Int}_{(X,d)}(A) = \bigcup\{U : U \subseteq A \text{ and } U \text{ is } d\text{-open in } X\}.$$

Remarks

1. When it is clear from the context, we usually just write $\mathrm{Int}(A)$ and suppress mention of the metric space (X,d).
2. The interior of a set is always an open set.
3. It is always the case that $\mathrm{Int}(A) \subseteq A$.

Worked Exercise 4.2

Find the interior of $(0,1) \cup \{2\}$ for the Euclidean metric on $\mathbb{R}$.

Solution

Let $A = (0,1) \cup \{2\}$.

Since $\text{Int}(A) \subseteq A$ and $(0,1)$ is an open set contained in A, we know that

$$(0,1) \subseteq \text{Int}(A) \subseteq (0,1) \cup \{2\}.$$

Thus we need only determine whether 2 is in the interior of A. However, since $\text{Int}(A)$ is open, Theorem 2.9 tells us that $x \in \text{Int}(A)$ if, and only if, there is $r > 0$ so that $B(x,r) \subseteq \text{Int}(A)$. But for each $r > 0$, $2 + \frac{1}{2}r \in B(2,r)$ and $2 + \frac{1}{2}r \notin A$ and so cannot be in $\text{Int}(A)$. Hence $2 \notin \text{Int}(A)$ and we conclude that

$$\text{Int}((0,1) \cup \{2\}) = (0,1).$$

Exercise 4.1

Let (X,d) be a metric space. Find the interior (in (X,d)) of $\emptyset$ and X.

Exercise 4.2

Let (X,d) be a metric space. Show that if A and B are subsets of X with $A \subseteq B$, then $\text{Int}(A) \subseteq \text{Int}(B)$.

Theorem 4.3

Let (X,d) be a metric space and let $A, B \subseteq X$. Then:

(a) $\text{Int}_{(X,d)}(A) \cup \text{Int}_{(X,d)}(B) \subseteq \text{Int}_{(X,d)}(A \cup B)$

(b) $\text{Int}_{(X,d)}(A) \cap \text{Int}_{(X,d)}(B) = \text{Int}_{(X,d)}(A \cap B)$.

Proof We use the result of Exercise 4.2.

(a) Since both $A \subseteq A \cup B$ and $B \subseteq A \cup B$, it follows from Exercise 4.2 that $\text{Int}(A) \subseteq \text{Int}(A \cup B)$ and $\text{Int}(B) \subseteq \text{Int}(A \cup B)$. Hence

$$\text{Int}(A) \cup \text{Int}(B) \subseteq \text{Int}(A \cup B).$$

(b) Similarly, since $A \cap B \subseteq A$ and $A \cap B \subseteq B$, we obtain

$$\text{Int}(A \cap B) \subseteq \text{Int}(A) \cap \text{Int}(B).$$

$\text{Int}(A) \cap \text{Int}(B)$ is the intersection of two open sets and hence is open.

Moreover, $\text{Int}(A) \subseteq A$ and $\text{Int}(B) \subseteq B$, so $\text{Int}(A) \cap \text{Int}(B) \subseteq A \cap B$. Since, by Definition 4.1, $\text{Int}(A \cap B)$ is the largest open subset of $A \cap B$, it follows that

$$\text{Int}(A) \cap \text{Int}(B) \subseteq \text{Int}(A \cap B).$$

Therefore

$$\text{Int}(A) \cap \text{Int}(B) = \text{Int}(A \cap B). \quad \blacksquare$$

Exercise 4.3

For $\mathbb{R}$ with the Euclidean metric, let $A = [0, 1]$ and $B = (1, 2]$. Show that $\text{Int}(A) \cup \text{Int}(B)$ is a *proper* subset of $\text{Int}(A \cup B)$.

The use of Theorem 2.9 in Worked Exercise 4.2 suggests that it is possible to characterise the interior of a set via a property of the points that lie in it (in a similar way to what we did with the closure) and this is indeed the case, as the following result shows.

Theorem 4.4

Let (X, d) be a metric space and let $A \subseteq X$.

The interior of A is given by

$$\text{Int}_{(X,d)}(A) = \{x \in X : \text{there is } r > 0 \text{ so that } B_d(x, r) \subseteq A\}.$$

Proof We first show that

$$V = \{x \in X : \text{there is } r > 0 \text{ so that } B_d(x, r) \subseteq A\}$$

is an open set contained within the interior of A.

For $x \in V$, let r_x be chosen so that $B_d(x, r_x) \subseteq A$; then V can be written as

$$V = \bigcup_{x \in V} B_d(x, r_x) \subseteq A.$$

In particular, V is the union of a collection of open sets, so it is itself open, and it is contained within A, hence it is a subset of $\text{Int}(A)$: $V \subseteq \text{Int}(A)$.

We now show that $\text{Int}(A) \subseteq V$: if $\text{Int}(A) - V \neq \emptyset$, then there is $x \in \text{Int}(A) - V$ and so, since $\text{Int}(A)$ is an open set, we can use Theorem 2.9 to find $r > 0$ for which $B_d(x, r) \subseteq \text{Int}(A)$. But $\text{Int}(A) \subseteq A$ and so $B_d(x, r) \subseteq A$, implying that $x \in V$, which is a contradiction. Hence $\text{Int}(A) \subseteq V$.

Combining these two set inclusions, we conclude that $\text{Int}(A) = V$, as required. $\blacksquare$

This characterisation is very useful when trying to find the interior of a given set and so we give a name to points that have the property described in the above theorem.

Definition 4.5 *Interior point*

Let (X, d) be a metric space and let $A \subseteq X$.

We say $x \in X$ is a **d-interior point of A** if there is an $r > 0$ so that $B_d(x, r) \subseteq A$.

Theorem 4.4 then tells us that

$$\text{Int}_{(X,d)}(A) = \{x \in X : x \text{ is a } d\text{-interior point of } A\}.$$

Exercise 4.4

Find the interior of $\mathbb{Q}$ for the usual Euclidean metric on $\mathbb{R}$.

The interior of a set A is open and, since open sets are the complements of closed sets, one would expect there is a characterisation of the interior of a set in terms of closed sets – and indeed there is. The characterisation is useful for converting results about closure to results about interior.

Theorem 4.6

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\text{Int}_{(X,d)}(A) = \left[\text{Cl}_{(X,d)}(A^c)\right]^c = X - \text{Cl}_{(X,d)}(X - A).$$

Proof We first show that $[\text{Cl}(A^c)]^c \subseteq \text{Int}(A)$.

Certainly, by the definition of closure, $A^c \subseteq \text{Cl}(A^c)$ and so it follows, on taking complements in X, that

$$[\text{Cl}(A^c)]^c \subseteq (A^c)^c = A.$$

But $[\text{Cl}(A^c)]^c$ is an open set, and so, by the definition of the interior, $[\text{Cl}(A^c)]^c \subseteq \text{Int}(A)$.

We now show that $\text{Int}(A) \subseteq [\text{Cl}(A^c)]^c$.

Since $\text{Int}(A) \subseteq A$, it follows that

$$A^c \subseteq [\text{Int}(A)]^c.$$

Moreover, since $\text{Int}(A)$ is open, $[\text{Int}(A)]^c$ is closed. Hence, we use Theorems 3.2 and 3.3 to find

$$\text{Cl}(A^c) \subseteq \text{Cl}([\text{Int}(A)]^c) = [\text{Int}(A)]^c.$$

So, on taking complements, we have

$$\text{Int}(A) \subseteq [\text{Cl}(A^c)]^c.$$

Combining these two set inclusions gives the required result. ■

Worked Exercise 4.7

Find the interior of $[0,1]$ for the Euclidean metric on $\mathbb{R}$.

Solution

By Theorem 4.6,

$$\text{Int}([0,1]) = [\text{Cl}([0,1]^c)]^c = [\text{Cl}((-\infty,0)\cup(1,\infty))]^c.$$

But $\text{Cl}((-\infty,0)\cup(1,\infty)) = (-\infty,0]\cup[1,\infty)$ and so

$$\text{Int}([0,1]) = ((-\infty,0]\cup[1,\infty))^c = (0,1).$$

Exercise 4.5

Let (X,d) be a metric space and $A \subseteq X$. Show that A is d-open if, and only if, $\text{Int}_{(X,d)}(A) = A$.

5 Boundaries of sets

In this section we introduce the notion of the boundary of a set.

We have already defined the interior and closure of a set A, and we have seen that $\text{Int}(A) \subseteq A \subseteq \text{Cl}(A)$. The points that lie in $\text{Cl}(A) - \text{Int}(A)$ are special. For example, in $\mathbb{R}$ with the Euclidean metric, the interior of $(0,1]$ is $(0,1)$ and the closure is $[0,1]$, and so $\text{Cl}((0,1]) - \text{Int}((0,1]) = \{0,1\}$. The points 0 and 1 are 'left over'; they seem to be acting as *boundary points*. This observation motivates the following definition.

Definition 5.1 *Boundary*

Let (X,d) be a metric space and let $A \subseteq X$.

The **d-boundary of A in X** is the set

$$\text{Bd}_{(X,d)}(A) = \text{Cl}_{(X,d)}(A) - \text{Int}_{(X,d)}(A).$$

Remarks

1. When it is clear from the context, we usually just write $\text{Bd}(A)$ and suppress mention of the metric space (X,d).

2. For any metric space (X, d), we can deduce directly from the definition of boundary that both $\mathrm{Bd}(\emptyset) = \emptyset$ and $\mathrm{Bd}(X) = \emptyset$. In particular, for $A \subseteq X$, $\mathrm{Bd}_{(X,d)}(A)$ need not be the same as $\mathrm{Bd}_{(A,d_A)}(A)$, which is always the empty set.

Our first result shows that the name 'boundary' is an appropriate one.

Theorem 5.2

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A) = \mathrm{Cl}_{(X,d)}(A) \cap \mathrm{Cl}_{(X,d)}(A^c).$$

In particular, $\mathrm{Bd}_{(X,d)}(A)$ is a closed set.

Proof This result follows from Theorem 4.6 in which we showed that $\mathrm{Int}(A) = \mathrm{Cl}(A^c)^c$, that is, $\mathrm{Int}(A) = X - \mathrm{Cl}(A^c)$. Hence

$$\begin{aligned}\mathrm{Bd}(A) &= \mathrm{Cl}(A) - \mathrm{Int}(A), \text{ by definition}\\ &= \mathrm{Cl}(A) \cap (\mathrm{Int}(A))^c\\ &= \mathrm{Cl}(A) \cap ((\mathrm{Cl}(A^c))^c)^c\\ &= \mathrm{Cl}(A) \cap \mathrm{Cl}(A^c),\end{aligned}$$

as claimed. ■

Exercise 5.1

Let (X, d) be a metric space and let $A \subseteq X$. Prove that

$$\mathrm{Bd}_{(X,d)}(A) = \mathrm{Bd}_{(X,d)}(A^c).$$

As with interior and closure, we can characterise the boundary of a set in terms of points with particular properties. These points are known as boundary points.

Definition 5.3 *Boundary point*

Let (X, d) be a metric space and let $A \subseteq X$.

A point $x \in X$ is a **d-boundary point of A** if for each $r > 0$, $B_d(x, r)$ meets both A and A^c (Figure 5.1).

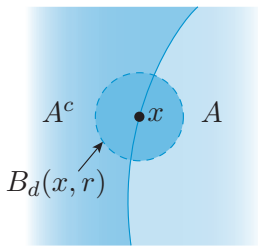


Figure 5.1 A boundary point of A

Theorem 5.4

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A) = \{x \in X : x \text{ is a } d\text{-boundary point of } A\}.$$

Proof Let $B = \{x \in X : x \text{ is a } d\text{-boundary point of } A\}$ – we must show that $\mathrm{Bd}(A) = B$.

We show that a point $b \in B$ if, and only if, $b \in \mathrm{Bd}(A)$.

A point $b \in B$ if, and only if, for each $r > 0$, $B(b, r) \cap A \neq \emptyset$ and $B(b, r) \cap A^c \neq \emptyset$.

But $B(b, r) \cap A \neq \emptyset$ for each $r > 0$ if, and only if, b is a closure point of A, and $B(b, r) \cap A^c \neq \emptyset$ for each $r > 0$ if, and only if, b is a closure point of A^c.

Now b is a closure point of A if, and only if, $b \in \mathrm{Cl}(A)$, by Theorem 3.8, and b is a closure point of A^c if, and only if, $b \in \mathrm{Cl}(A^c)$, also by Theorem 3.8.

Hence $b \in B$ if, and only if, $b \in \mathrm{Cl}(A) \cap \mathrm{Cl}(A^c)$. But Theorem 5.2 tells us that $\mathrm{Cl}(A) \cap \mathrm{Cl}(A^c) = \mathrm{Bd}(A)$ and the result follows. ■

Theorem 5.5

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A) = (\mathrm{Int}_{(X,d)}(A^c) \cup \mathrm{Int}_{(X,d)}(A))^c.$$

Proof We first show that $\mathrm{Bd}(A) \subseteq (\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c$ by using the characterisation of the boundary in terms of boundary points.

Let $x \in \mathrm{Bd}(A)$. Then x is not an interior point of A^c since $B_d(x, r) \cap A \neq \emptyset$ for each $r > 0$, and x is not an interior point of A since $B_d(x, r) \cap A^c \neq \emptyset$ for each $r > 0$. So

$$x \in \mathrm{Int}(A^c)^c \cap \mathrm{Int}(A)^c = (\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c,$$

by De Morgan's First Law. Thus

$$\mathrm{Bd}(A) \subseteq (\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c.$$

It remains only to verify that $(\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c \subseteq \mathrm{Bd}(A)$.

Let $x \in (\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c$. Then x is not an interior point of A, and so $B_d(x, r)$ intersects A^c for each $r > 0$. Similarly, x is not an interior point of A^c, and so $B_d(x, r)$ intersects A for each $r > 0$. Hence $B_d(x, r)$ intersects both A and A^c for each $r > 0$. So $x \in \mathrm{Bd}(A)$. Thus

$$(\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c \subseteq \mathrm{Bd}(A).$$

Therefore $\mathrm{Bd}(A) = (\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c$. ■

Remarks

1. Theorem 5.5 tells us immediately that $\mathrm{Bd}(A) \cup \mathrm{Int}(A^c) \cup \mathrm{Int}(A) = X$.
2. We deduce from Theorem 5.5 that $\mathrm{Bd}(A) \cap \mathrm{Int}(A^c) = \emptyset$ and $\mathrm{Bd}(A) \cap \mathrm{Int}(A) = \emptyset$. We saw earlier that $\mathrm{Int}(A) \cap \mathrm{Int}(A^c) = \emptyset$. Thus $\mathrm{Bd}(A)$, $\mathrm{Int}(A^c)$ and $\mathrm{Int}(A)$ are mutually disjoint.

The relationships between the interior and boundary of a set and its complement are illustrated in Figure 5.2. Note that $\mathrm{Bd}(A)$ is a closed set, disjoint from $\mathrm{Int}(A)$ and $\mathrm{Int}(A^c)$. It is difficult to do justice to this in a drawing.

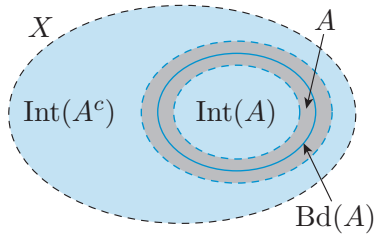


Figure 5.2 $\mathrm{Int}(A)$, $\mathrm{Int}(A^c)$ and $\mathrm{Bd}(A)$

Exercise 5.2

Let $\mathbb{R}$ have its Euclidean metric. Find the interior and boundary of (a) $\mathbb{Q}$ and (b) $\mathbb{Q}^c$.

Hint: use the result of Worked Exercise 3.9.

You are asked to consider the discrete metric in the next exercise.

Exercise 5.3

Let (X, d) be a metric space with the discrete metric and let A be a non-empty proper subset. Find the interior and boundary of A.

5.1 The algebra of boundaries

We turn now to some general results about the behaviour of interiors, exteriors and boundaries under unions and intersections.

Theorem 5.6

Let (X, d) be a metric space and let $A, B \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A \cup B) \subseteq \mathrm{Bd}_{(X,d)}(A) \cup \mathrm{Bd}_{(X,d)}(B).$$

Proof We use the characterisation of the boundary in terms of closure, given in Theorem 5.2, together with the behaviour of closures under unions and intersections given in Theorem 3.5:

We can also prove this result by using the results of Theorem 4.3.

$$\begin{aligned}
\mathrm{Bd}(A \cup B) &= \mathrm{Cl}(A \cup B) \cap \mathrm{Cl}((A \cup B)^c), && \text{by Theorem 5.2} \\
&= \mathrm{Cl}(A \cup B) \cap \mathrm{Cl}(A^c \cap B^c), && \text{by De Morgan's First Law} \\
&\subseteq (\mathrm{Cl}(A) \cup \mathrm{Cl}(B)) \cap (\mathrm{Cl}(A^c) \cap \mathrm{Cl}(B^c)), && \text{by Theorem 3.5} \\
&= (\mathrm{Cl}(A) \cap \mathrm{Cl}(A^c) \cap \mathrm{Cl}(B^c)) \cup (\mathrm{Cl}(B) \cap \mathrm{Cl}(A^c) \cap \mathrm{Cl}(B^c)) \\
&\subseteq (\mathrm{Cl}(A) \cap \mathrm{Cl}(A^c)) \cup (\mathrm{Cl}(B) \cap \mathrm{Cl}(B^c)) \\
&= \mathrm{Bd}(A) \cup \mathrm{Bd}(B), && \text{by Theorem 5.2.} \quad \blacksquare
\end{aligned}$$

Here we are using the identity $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$, which is a special case of the more general identity $(\cup_{i \in I} E_i) \cap G = \bigcup_{i \in I} (E_i \cap G)$ and is an example of a *distributive law*.

Exercise 5.4

Show that for all sets E, F, G the identity $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ holds.

Exercise 5.5

Let (X, d) be a metric space and let $A, B \subseteq X$. Use Theorem 5.6 and the result of Exercise 5.1 to show that

$$\mathrm{Bd}_{(X,d)}(A \cap B) \subseteq \mathrm{Bd}_{(X,d)}(A) \cup \mathrm{Bd}_{(X,d)}(B).$$

6 The size of sets

In this section we look at two different ways of describing when sets are big or small. The first approach is metrical in nature, in that it measures the size of a set in terms of the metric of the underlying space. The second approach is a purely 'set theoretic' one that extends the familiar idea of saying that a finite set has n elements to the situation when the set is not finite. It does not depend on any notion of distance and has some surprising consequences: there are different 'sizes' of infinite sets.

6.1 Dense and nowhere dense sets

In this short subsection, we investigate two special properties that a subset of the underlying set X of a metric space may possess. These properties give an idea of the 'size' of a set but from a very particular viewpoint – that of the metric. The indicators of size that we will obtain are very crude and only allow for three possible values: 'small', 'big' and 'somewhere in between'. Despite this, the metrical approach has turned out to be useful in modern mathematics, mainly because the notion of 'big' that we will shortly define has some nice properties: the intersection of two 'big' open sets is itself a 'big' open set. (In fact it turns out that in many metric spaces the intersection of a countable collection of 'big' open sets is 'big' (but not necessarily open). This result, known as the Baire Category Theorem, has played a fundamental role in the development of modern mathematics, particularly in functional analysis.)

In Worked Exercise 3.9 we used the fact that there is a rational number between any two distinct real numbers – that is, that the rational numbers are *dense* in $\mathbb{R}$ – to help us show that if $\mathbb{R}$ carries its Euclidean metric, then $\mathrm{Cl}(\mathbb{Q}) = \mathbb{R}$. In contrast, if $\mathbb{R}$ has its discrete metric, then the closure of every set is equal to the set itself. In particular, even an apparently very large set, such as $\mathbb{Q}$ or $\mathbb{R} - \mathbb{Q}$, would have closure not equal to $\mathbb{R}$. In fact, for the discrete metric, the only subset of $\mathbb{R}$ whose closure is $\mathbb{R}$ is $\mathbb{R}$ itself. A subset of $\mathbb{R}$ whose closure (with respect to a particular metric) is equal to the whole of $\mathbb{R}$ is called a *dense* subset of $\mathbb{R}$. More generally, we make the following definitions.

The closure of subsets for the discrete metric was discussed in Worked Exercise 1.11.

Definition 6.1 *Dense and nowhere dense*

Let (X, d) be a metric space and let $A \subseteq X$.

The set A is **d-dense in** X if $\mathrm{Cl}_{(X,d)}(A) = X$.

The set A is **nowhere dense in** X if $\mathrm{Int}_{(X,d)}(\mathrm{Cl}_{(X,d)}(A)) = \emptyset$.

Remarks

1. The notion of dense corresponds to 'big' and nowhere dense corresponds to 'small'.
2. Since $\mathrm{Cl}(X) = X$, the underlying set X of a metric space is always dense in itself.
3. For any metric space, since $\mathrm{Int}(\mathrm{Cl}(\emptyset)) = \mathrm{Int}(\emptyset) = \emptyset$, the empty set is nowhere dense.
4. A set A can be neither dense nor nowhere dense. Consider the interval $[0, 1]$ in $\mathbb{R}$ with the usual Euclidean metric: $\mathrm{Cl}([0, 1]) = [0, 1] \neq \mathbb{R}$ and

   $$\mathrm{Int}(\mathrm{Cl}([0, 1])) = \mathrm{Int}([0, 1]) = (0, 1) \neq \emptyset.$$

   Most sets are of this type – 'somewhere in between'.

Determining whether a given set is dense (or nowhere dense) in a space (X, d) is an exercise in the calculation of closures and interiors. In Worked Exercise 3.9 we saw that $\mathrm{Cl}(\mathbb{Q}) = \mathbb{R}$ for the usual Euclidean metric and so $\mathbb{Q}$ is dense in $\mathbb{R}$ (for the Euclidean metric).

Worked Exercise 6.2

Let (X, d) be a metric space and let $A \subseteq X$ be dense in X. Show that if $A \subseteq B$ then B is also dense in X.

Solution

If $A \subseteq B$ then, by Theorem 3.2, $\mathrm{Cl}(A) \subseteq \mathrm{Cl}(B)$. But A is dense in X, so $\mathrm{Cl}(A) = X$. Hence $X = \mathrm{Cl}(A) \subseteq \mathrm{Cl}(B) \subseteq X$ and so $\mathrm{Cl}(B) = X$. That is, B is dense in X.

Exercise 6.1

Let (X, d) be a metric space and let $A \subseteq X$ be nowhere dense in X. Show that if $B \subseteq A$, then B is nowhere dense in X.

Exercise 6.2

Let (X, d) be a metric space and let U and V be d-open subsets of X that are d-dense in X.

Show that $U \cap V$ is also (a d-open subset of X that is) d-dense in X.

Dense sets and subspaces

If (A, d_A) is a metric subspace of the metric space (X, d) and we wish to find sets that are dense in A, then we need to take special care to track which metric we are calculating denseness with regard to. There is, however, a theorem that relates these two closures.

Theorem 6.3

Let (X, d) be a metric and let $A \subset X$. Let d_A denote the subspace metric for A.

A set B contained in A is d_A-dense in A if, and only if, $A \subseteq \mathrm{Cl}_{(X,d)}(B)$.

Proof This is a proof where it is important to track which metric we are using at each stage.

We need to show that $\mathrm{Cl}_{(A,d_A)}(B) = A$ if, and only if, $A \subseteq \mathrm{Cl}_{(X,d)}(B)$.

We show that $\mathrm{Cl}_{(X,d)}(B) \supseteq A$ implies $\mathrm{Cl}_{(A,d_A)}(B) = A$ by proving the contrapositive: $\mathrm{Cl}_{(A,d_A)}(B) \neq A$ implies $\mathrm{Cl}_{(X,d)}(B) \not\supseteq A$.

It is certainly the case that $\mathrm{Cl}_{(A,d_A)}(B) \subseteq A$, so suppose $a \in A - \mathrm{Cl}_{(A,d_A)}(B)$; then a is not a d_A-closure point of B and we can find $r > 0$ for which $B_{d_A}(a, r) \cap B = \emptyset$. That is,

$$\{x \in A : d_A(a, x) < r\} \cap B = \emptyset.$$

This set equality can be written as

$$B_d(a, r) \cap A \cap B = \emptyset.$$

But $B \subseteq A$ and so this is the same as

$$B_d(a, r) \cap B = \emptyset,$$

and so $a \notin \mathrm{Cl}_{(X,d)}(B)$. Hence $A \not\subseteq \mathrm{Cl}_{(X,d)}(B)$.

Now suppose that $\mathrm{Cl}_{(A,d_A)}(B) = A$ – we aim to show that $\mathrm{Cl}_{(X,d)}(B)$ contains A. So let $a \in A$ and $r > 0$. Since $\mathrm{Cl}_{(A,d_A)}(B) = A$, there is $b \in B$ with $d_A(a, b) < r$; but $d_A(a, b) = d(a, b)$ and so we deduce that $B_d(a, r)$ also meets B. Hence $a \in \mathrm{Cl}_{(X,d)}(B)$ and we conclude that $A \subseteq \mathbb{C}_{(X,d)}(B)$, as required. ■

This result lets us turn questions about dense sets in subspaces into questions about closures of sets in the original space.

6.2 Countable and uncountable sets

In Worked Exercise 3.9 we saw that $\mathrm{Cl}(\mathbb{Q}) = \mathbb{R}$ for the usual Euclidean metric on $\mathbb{R}$ and so $\mathbb{Q}$ is a dense subset of $\mathbb{R}$. This may not strike you as a particularly surprising result, but in this subsection we show that although $\mathbb{Q}$ is an infinite set, it is not as big as $\mathbb{R}$, having strictly fewer points. The results in this subsection may strike you as rather surprising and counterintuitive – after all, how can you have different sizes of infinity? But they are correct and their implications are profound.

We start by recalling that a *finite set* is one with finitely many elements, and that an *infinite set* is one that is not finite. In this subsection we restrict ourselves to categorising infinite sets into two types – those that are *countable* and those that are *uncountable*. Informally, we say that an infinite set is 'countable' if all its elements can be written down as an infinite 'list', one after another, so that they can then be *counted*. An infinite set that cannot be listed in this way is said to be 'uncountable'. (It is not immediately apparent that there are any sets that are uncountable and one of the main aims of this subsection is to show that some sets with which you are already familiar are, in fact, uncountable.)

The empty set, having no elements, is regarded as being a finite set.

In fact, using similar techniques to those described here, one can show that there are different types of uncountable set and that there is a whole hierarchy of different 'sizes' of infinity. We will not do that here.

For example, the elements of the set $\mathbb{N}$ of natural numbers can be listed in the following (obvious!) way:

$$1, 2, 3, \ldots;$$

so this set is countable (according to our informal definition). Similarly, the set $\mathbb{Z}$ of integers can be listed as

$$0, 1, -1, 2, -2, \ldots$$

Every integer appears somewhere in the list.

and so this set too is countable (according to our informal definition).

Note that if the elements of a set can be *listed*, then they can be *numbered*. For example, using our listing for the set $\mathbb{Z}$, we can say that 0 is the first term in the listing of $\mathbb{Z}$, 1 is the second term, -1 is the third term, and so on. Thus we can assign a natural number to each of them, as follows:

$$0 \mapsto 1,\ 1 \mapsto 2,\ -1 \mapsto 3,\ \ldots.$$

In effect, we have defined a one–one map from $\mathbb{Z}$ to $\mathbb{N}$.

However, it is not obvious whether we could list all the elements of the set $\mathbb{R}$ of real numbers.

These observations lead to the following definition.

Definition 6.4 *Countable and uncountable*

A set A is **countable** if there exists a one–one map $f\colon A \to \mathbb{N}$, and is **uncountable** otherwise.

A countable set with infinitely many elements is said to be **countably infinite**.

Some authors call a set A countable if there is a one–one *and onto* map from A to $\mathbb{N}$ – this alternative definition does *not* classify finite sets as countable.

Remarks

1. Since the definition does not require the one–one map to be *onto*, it follows that all finite sets (including $\emptyset$) are countable. For example, $\{7, \pi, \sqrt{2}, -31, \pi\sqrt{2}\}$ can be mapped one–one to $\{1, 2, 3, 4, 5\}$, a finite subset of $\mathbb{N}$, and so is countable.
2. If a set A is countable, so that there exists a one–one map from A to $\mathbb{N}$, then we can *list* its elements in the order defined by this one–one map. Conversely, if we can list the elements of a set A, then the ordering of the list defines a one–one map from A to $\mathbb{N}$. Hence a set is countable if, and only if, it is *listable*.
3. Every subset of a countable set is countable and this is shown in the worked exercise following these remarks. However, a subset of an uncountable set may be countable (finite or infinite) or uncountable.
4. The main goal of this subsection is to show you that there are some infinite sets that are uncountable. Until then, you will have to take it on trust that this is a sensible definition to make.

Worked Exercise 6.5

Suppose that $A \subseteq B$ and B is countable. Show that A is also countable.

Solution

Let $f\colon B \to \mathbb{N}$ be a one–one map. Such a map exists since we are assuming B is countable.

Define $f|_A\colon A \to \mathbb{N}$ by $f_A(a) = f(a)$ for each $a \in A$, so $f|_A$ is the restriction of f to the set A. Then $f|_A$ is also one–one, for if $f|_A(a) = f|_A(b)$, then $f(a) = f(b)$ and so $a = b$, since f is one–one. We conclude that A is countable.

We now investigate the countability of some of the standard subsets of $\mathbb{R}$. We begin by noting that the set $\mathbb{N}$ is clearly countable, since the map $f\colon \mathbb{N} \to \mathbb{N}$ defined by $f(n) = n$ is a one–one map from $\mathbb{N}$ to $\mathbb{N}$.

Exercise 6.3

Show that the set $\mathbb{Z}$ of integers is countable.

You may find the next result surprising.

Theorem 6.6

The set $\mathbb{Q}$ of rational numbers is countable.

Proof There are several ways of proving this result. Recall that any non-zero rational number can be written uniquely as $\frac{p}{q}$, where $p \in \mathbb{Z}$ and $q \in \mathbb{N}$ have no common factors. If we agree to give 0 the representation $\frac{0}{1}$, we have a unique representation of the form $\frac{p}{q}$ for each element of $\mathbb{Q}$. We now construct a one–one map from $\mathbb{Q}$ into $\mathbb{Z}$ as follows:

$$f : \frac{p}{q} \mapsto \begin{cases} 2^p 3^q, & \text{if } p \geq 0, \\ -2^{-p} 3^q, & \text{if } p < 0. \end{cases}$$

For example, $f(\frac{3}{4}) = 2^3\,3^4 = 648$ and $f(-\frac{3}{4}) = -2^{-(-3)}\,3^4 = -648$.

We know that $\mathbb{Z}$ is countable (by Exercise 6.3) and so there exists a one–one map from $\mathbb{Z}$ to $\mathbb{N}$. Composing these two maps, we obtain a one–one map from $\mathbb{Q}$ to $\mathbb{N}$. Thus $\mathbb{Q}$ is countable. ■

So far, we have looked only at countable subsets of $\mathbb{R}$, and you may be wondering whether *all* subsets of $\mathbb{R}$ are countable. In fact, there are infinitely many *un*countable subsets of $\mathbb{R}$. We now show that the subset of $\mathbb{R}$ comprising the open interval $(0, 1)$ is uncountable.

Theorem 6.7

The set $(0, 1)$ is uncountable.

Proof We prove this by supposing that $(0, 1)$ is countable and obtaining a contradiction.

If $(0, 1)$ is countable, then we can list all its elements in decimal form:

$$\begin{aligned} x_1 &= 0.\boxed{d_{1,1}}\, d_{1,2}\, d_{1,3} \ldots \\ x_2 &= 0.\, d_{2,1} \boxed{d_{2,2}}\, d_{2,3} \ldots \\ x_3 &= 0.\, d_{3,1}\, d_{3,2} \boxed{d_{3,3}} \ldots \\ &\vdots \quad \vdots \quad \vdots \end{aligned}$$

where each $d_{i,j}$ belongs to $\{0, 1, \ldots, 9\}$.

Terminating decimals such as 0.1 have two different decimal expansions: either 0.1000 or $0.0999\ldots$. In this case, we only use the 'terminating' expression that ends in an infinite number of 0s.

We now choose a real number

$$x = 0.a_1 a_2 a_3 \ldots$$

in such a way that

$$a_1 \neq d_{1,1},\ a_2 \neq d_{2,2},\ a_3 \neq d_{3,3},\ \ldots \text{ and } a_i \neq 9 \text{ for all } i.$$

By ensuring that $a_i \neq 9$ for all i, we have made certain that x does not end in an infinite sequence of 9s, and so is not equivalent to a number in our list ending in an infinite sequence of 0s.

For example, we could define a_n for $n \in \mathbb{N}$ by

$$a_n = \begin{cases} 0, & \text{if } d_{n,n} \neq 0, \\ 1, & \text{if } d_{n,n} = 0. \end{cases}$$

We have now obtained a real number x in $(0, 1)$ that does not appear in our list, since it differs from x_n at the nth decimal place. This, however, is a contradiction since the list was supposed to contain *all* the numbers in $(0, 1)$. Thus our original supposition that $(0, 1)$ is countable must be wrong, and so $(0, 1)$ is uncountable. ■

Corollary 6.8

The set $\mathbb{R}$ is uncountable.

Proof If $\mathbb{R}$ were countable, then the subset $(0, 1)$ would be countable. This contradicts Theorem 6.7, and hence $\mathbb{R}$ is uncountable. ■

Exercise 6.4

Show that any open interval (a, b) with $a < b$ is uncountable.

Hint: find a one–one map from $(0, 1)$ to (a, b).

By using the results of Worked Exercise 6.5 and Exercise 6.4, we can deduce that open intervals of the form (a, ∞) and $(-\infty, b)$ are uncountable, since each contains an uncountable subset (a, b) for $a < b$. (If (a, ∞) were countable, then Worked Exercise 6.5 implies that any subset of it would also be countable and we have just seen in Exercise 6.4 that (a, b) is uncountable whenever $a < b$.)

When dealing with metric spaces, we occasionally need to consider the union of two or more sets. In fact (perhaps surprisingly), the union of *countably* many countable sets is always countable.

Theorem 6.9

For each $n \in \mathbb{N}$, let A_n be a countable set. Then $A = \bigcup_{n=1}^{\infty} A_n$ is also countable.

Proof Since each set A_n is countable, we can list the elements of A_1 as

$$x_{1,1}, x_{1,2}, x_{1,3}, \ldots;$$

the elements of A_2 as

$$x_{2,1}, x_{2,2}, x_{2,3}, \ldots;$$

and so on. We can then list all the elements of A as shown below.

We start our list with the element $x_{1,1}$, at the top left of the array (whose subscripts sum to 2). Next, we take the elements $x_{1,2}$ and $x_{2,1}$ on the diagonal immediately to the right of $x_{1,1}$ (whose subscripts sum to 3), then the elements on the next diagonal (whose subscripts sum to 4), and so on.

$$\begin{array}{ccccccccccc}
x_{1,1} & \rightarrow & x_{1,2} & & x_{1,3} & \rightarrow & x_{1,4} & & x_{1,5} & \cdots \\
 & \swarrow & & \nearrow & & \swarrow & & \nearrow & & \\
x_{2,1} & & x_{2,2} & & x_{2,3} & & x_{2,4} & & \cdot & \cdots \\
\downarrow & \nearrow & & \swarrow & & \nearrow & & & & \\
x_{3,1} & & x_{3,2} & & x_{3,3} & & x_{3,4} & & \cdot & \cdots \\
 & \swarrow & & \nearrow & & & & & & \\
x_{4,1} & & x_{4,2} & & x_{4,3} & & x_{4,4} & & \cdot & \cdots \\
\downarrow & \nearrow & & & & & & & & \\
x_{5,1} & & \cdot & & \cdot & & \cdot & & \cdot & \cdots \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots &
\end{array}$$

This gives us a list starting $x_{1,1}, x_{1,2}, x_{2,1}, x_{3,1}, x_{2,2}, x_{1,3}, x_{1,4}, \ldots$.

Every element $x_{i,j}$ appears somewhere in the list.

Thus A is listable, and therefore countable. ■

There is an obvious corollary.

Corollary 6.10

The union of finitely many countable sets is countable.

Worked Exercise 6.11

Show that $\mathbb{Q} \times \mathbb{Q}$ is a countable subset of $\mathbb{R}^2$.

Solution

It suffices to observe that

$$\mathbb{Q} \times \mathbb{Q} = \bigcup_{p \in \mathbb{Q}} \{p\} \times \mathbb{Q}$$

and so $\mathbb{Q} \times \mathbb{Q}$ can be written as the countable union of countable sets. Hence Theorem 6.9 implies that $\mathbb{Q} \times \mathbb{Q}$ is countable.

7 Continuity and open and closed sets

In this short section we look briefly at how some of the ideas we have examined in this book may be extended further. In particular we look at the relationship between continuous functions and open and closed sets, and we obtain a characterisation of continuous functions between metric spaces in terms of both open and closed sets. This characterisation turns out to be the starting point for the theory of *point-set topology*; although this is outside the scope of this module, we give a brief overview of it.

This section is not assessed.

Theorem 7.1

Let (X, d) and (Y, e) be metric spaces and suppose that $f : X \to Y$. Then f is (d, e)-continuous on X if, and only if, for each e-closed set $E \subseteq Y$, $f^{-1}(E)$ is d-closed.

Proof *If:*

Suppose that f is (d, e)-continuous on X and let E be an e-closed subset of Y. We must show that $f^{-1}(E) = \{x \in X : f(x) \in E\}$ is a d-closed subset of X. So let (x_n) be a d-convergent sequence in X with limit x and suppose that $x_n \in f^{-1}(E)$ for each n. Our aim is to show that $x \in f^{-1}(E)$.

Then $(f(x_n))$ is a sequence in Y for which $f(x_n) \in E$ for each $n \in \mathbb{N}$. Moreover, since f is (d, e)-continuous, the sequence $(f(x_n))$ is (d, e)-convergent with limit $f(x)$. But E is e-closed and so $f(x) \in E$, implying that $x \in f^{-1}(E)$, as required.

Only if:

Suppose that if $E \subseteq Y$ is an e-closed set, then $f^{-1}(E)$ is d-closed. We must show that f is (d, e)-continuous; that is, for each d-convergent sequence (x_n) in X with limit x, say, we have $(f(x_n))$ is e-convergent with limit $f(x)$.

So suppose that (x_n) is a d-convergent sequence in X with limit x and yet $f(x_n)$ does not e-converge to $f(x)$. Then there is $\varepsilon > 0$ such that for infinitely many n, $f(x_n) \notin B_e(f(x), \varepsilon)$. By restricting attention to a subsequence of (x_n) if necessary, we may assume that for each n, $f(x_n) \notin B_e(f(x), \varepsilon)$. Now $B_e(f(x), \varepsilon)$ is an e-open set and so $Y - B_e(f(x), \varepsilon)$ is e-closed and $f(x_n) \in Y - B_e(f(x), \varepsilon)$ for each n. By hypothesis, $f^{-1}(Y - B_e(f(x), \varepsilon))$ is d-closed and we also have $x_n \in f^{-1}(Y - B_e(f(x), \varepsilon))$ by definition. We assumed that (x_n) was a convergent sequence and so the limit of (x_n), the point x, must lie in $f^{-1}(Y - B_e(f(x), \varepsilon))$, meaning that $f(x) \in Y - B_e(f(x), \varepsilon)$, which is impossible. So no such sequence (x_n) can exist and f is (d, e)-continuous as required. ■

Since open sets are the complement of closed sets in a metric space, we expect a version of this result in terms of open sets – and indeed, this is the case.

Corollary 7.2

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$. Then f is (d, e)-continuous on X if, and only if, for each e-open set $U \subseteq Y$, $f^{-1}(U)$ is d-open.

In order to prove this corollary, we need to know how complements of sets interact with map inverses.

Exercise 7.1

Let $f\colon X \to Y$ and $A \subseteq Y$. Show that

$$f^{-1}(Y - A) = X - f^{-1}(A).$$

We can now prove the corollary.

Proof of Corollary 7.2 Let U be an e-open set.

Then $Y - U$ is e-closed and so $f^{-1}(Y - U) = X - f^{-1}(U)$ is d-closed if, and only if, f is (d, e)-continuous.

But $X - f^{-1}(U)$ is d-closed if, and only if, $f^{-1}(U)$ is d-open.

Hence f is (d, e)-continuous if, and only if, for each e-open set U, $f^{-1}(U)$ is d-open. ■

Drawing together the results we have proved in this chapter, we make the following observations.

If (X, d) is a metric space, then the collection, $\mathcal{T}$ say, of *d-open sets* has the following properties:

(T1) Both $\emptyset$ and X are in $\mathcal{T}$.

(T2) If U and V are in $\mathcal{T}$, then so is $U \cap V$.

(T3) If $\{U_i : i \in I\}$ is a collection of sets that are all in $\mathcal{T}$, then

$$\bigcup_{i \in I} U_i \text{ is in } \mathcal{T}.$$

In addition, if $f\colon X \to Y$ and (Y, e) is a metric space, then f is (d, e)-continuous on X if, and only if, for each e-open set U in Y, $f^{-1}(U)$ is d-open in X.

These observations are the basis for a significant extension of the definition of continuity.

Definition 7.3 *Topology*

We define a **topology** for a set X to be any collection $\mathcal{T}$ of subsets of X that satisfy (T1), (T2) and (T3). We call the sets that are in $\mathcal{T}$ the *open sets*.

A *topological space* is a set X together with a collection of subsets of X, $\mathcal{T}$ say, that form a topology for X, denoted $(X, \mathcal{T})$.

Remarks

1. We have seen in this chapter how metric spaces give rise to a collection of sets that satisfy these conditions. A topology on a space X that is formed from a metric d in this way is called a *metrisable topology* and when we wish to emphasise the dependence on the underlying metric, we write $\mathcal{T}(d)$.

2. There are topologies that are not metrisable – that is, they are not the open sets that are associated with a metric. For a simple example of such a topology, consider $(\mathbb{R}, \mathcal{T})$ where $\mathcal{T} = \{\emptyset, \mathbb{R}\}$. It is easy to verify that $\mathcal{T}$ satisfies (T1)–(T3) and so defines a topology for $\mathbb{R}$. However, it is not one determined by a metric, since – by Lemma 1.2 – single points are closed sets for any metric d on $\mathbb{R}$; and so in particular the complement of a single point, such as $\mathbb{R} - \{0\}$, would be d-open. But $\mathbb{R} - \{0\}$ is not in $\mathcal{T}$ and so $\mathcal{T}$ cannot be generated from the metric – it is not a metrisable topology.

Moreover, in view of Corollary 7.2, given two topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ we can define what it means for a function to be continuous between them in a way that is consistent with the definition of continuity for metric spaces.

Definition 7.4 *Continuity in a topological space*

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces and let $f\colon X \to Y$. We say that f is $(\mathcal{T}_X, \mathcal{T}_Y)$-continuous if for each $U \in \mathcal{T}_Y$, $f^{-1}(U) \in \mathcal{T}_X$.

The main advantage of topologies is that there are spaces that are so 'large' it is not possible to define a metric for them that captures the properties that we may be interested in, and yet it is possible to define topologies for them that do.

Learning outcomes

We list those topics on which we may set assessment questions to test your understanding of the chapter.

After working through this chapter you should be able to:

(a) State and use the definition of a *closed set*.

(b) Determine whether a given set in a metric space is closed.

(c) State and use the definition of an *open set*.

(d) Determine whether a given set in a metric space is open.

(e) Describe and use the basic properties of open and closed sets.

(f) Use various results concerning the union and intersection of sets, including De Morgan's Laws.

(g) Find the closure points and closure of a given set.

(h) Find the interior points and interior of a given set.

(i) Find the boundary points and boundary of a given set.

(j) Describe and use the basic properties of closures, interiors and boundaries.

(k) Determine whether a given set is dense or nowhere dense.

(l) Distinguish between finite, countably infinite and uncountable sets.

Solutions to exercises

Solution to Exercise 1.1

(a) (i) The limit of the sequence, 0, is in A.

(ii) The limit of the sequence, 0, is not in A.

(b) (i) The limit of the sequence, $(1, 2)$, is in A.

(ii) The limit of the sequence, $(1, 2)$, is not in A.

Solution to Exercise 1.2

The terms of the sequence (x_n), where $x_n = a - \frac{1}{n}$, all lie in $(-\infty, a)$. However, the limit of the sequence (with respect to the Euclidean metric d) is a, which does not lie in $(-\infty, a)$. Hence $(-\infty, a)$ is not closed for the Euclidean metric.

Solution to Exercise 1.3

(a) $A^c = \{2\}$; (b) $A^c = \{2, 4, 5\}$.

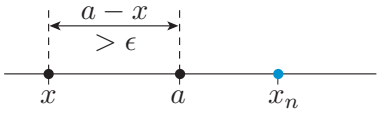
Solution to Exercise 1.4

$A^c = \mathbb{R} - [0, 1] = (-\infty, 0) \cup (1, \infty)$.

Solution to Exercise 1.5

The complement of $(-\infty, a)$ in $\mathbb{R}$ is $[a, \infty)$ and the argument to show that this set is d-closed mirrors that of Worked Exercise 1.4.

Let (x_n) be a d-convergent sequence in $\mathbb{R}$ with limit x, and suppose that $x_n \in [a, \infty)$ for each $n \in \mathbb{N}$. In order to show that the interval $[a, \infty)$ is d-closed we need to show that $x \in [a, \infty)$. Thus our goal is to show that $x \geq a$.



So suppose instead that $x < a$; then $|x - x_n| = x_n - x \geq a - x > 0$ for each $n \in \mathbb{N}$. Thus if we choose $\varepsilon > 0$ for which $a - x > \varepsilon$, as shown in the figure, then we have $|x - x_n| > \varepsilon$ for each $n \in \mathbb{N}$, contradicting the fact that $x_n \to x$ as $n \to \infty$. Hence $x \geq a$ and $[a, \infty)$ is d-closed as required.

Solution to Exercise 1.6

Now

$$\begin{aligned} A &= \{y \in X : d(x, y) \geq r\} \\ &= X - \{y \in X : d(x, y) < r\} \\ &= X - B_d(x, r). \end{aligned}$$

Hence, by the result of Worked Exercise 1.6, A is d-closed.

Solution to Exercise 1.7

Let $a = \lim_{n\to\infty} a_n$ and suppose that $a < m$. Then there is $\varepsilon > 0$ such that $a < m - \varepsilon$. Hence, we can calculate that for each $n \in \mathbb{N}$,

$$a < m - \varepsilon \leq a_n - \varepsilon$$

and so rearranging gives

$$\varepsilon < a_n - a \text{ for each } n \in \mathbb{N}.$$

This means that $(|a_n - a|)$ is not a null sequence and so $a_n \not\to a$ as $n \to \infty$, which is a contradiction. Thus $a \geq m$.

Similarly, if $M < a$, then there is $\varepsilon > 0$ such that $M + \varepsilon < a$. Hence, we can calculate that for each $n \in \mathbb{N}$,

$$a_n \leq M < M + \varepsilon < a$$

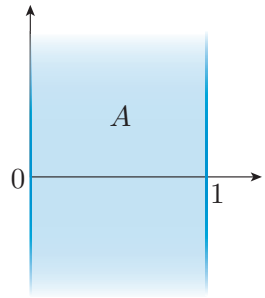
and rearranging gives

$$\varepsilon < a - a_n \text{ for each } n \in \mathbb{N}.$$

This means that $(|a_n - a|)$ is not a null sequence and so $a_n \not\to a$ as $n \to \infty$, which is a contradiction. Thus $a \leq M$.

Solution to Exercise 1.8

(a) A sketch is shown here. Note that the convention for boundaries is to draw them with solid lines if the boundary is included in the set and with dashed lines if it is not.



(b) Let $(\mathbf{a}_n = (a_{1,n}, a_{2,n}))$ be a convergent sequence in the plane, all of whose terms lie in A. So $(\mathbf{a}_n)$ has limit $\mathbf{a} = (a_1, a_2)$, say, and for each $n \in \mathbb{N}$, $0 \leq a_{1,n} \leq 1$.

Since $(\mathbf{a}_n)$ is convergent, we know from Theorem 3.15 of Chapter 15 that both of the real sequences $(a_{1,n})$ and $(a_{2,n})$ are also convergent with limits a_1 and a_2.

We now immediately use Lemma 1.8 to deduce that $0 \leq a_1 \leq 1$, implying that $\mathbf{a} \in A$. Hence A is closed for the Euclidean metric.

Solution to Exercise 1.9

Suppose that (f_n) is a $d_{\max}$-convergent sequence of functions in $C[0,1]$ with $f_n \in B_M$ for each n. We must show that the limit of the sequence, f, say, also lies in B_M. That is, we must show $f(0) = M$.

For each $n \in \mathbb{N}$ we have

$$|f(0) - M| = |f(0) - f_n(0)|$$

because $f_n \in B_M$. Hence

$$|f(0) - M| \leq d_{\max}(f_n, f),$$

which is a null sequence. We conclude that

$$|f(0) - M| \leq 0,$$

implying that $f(0) = M$ and so $f \in B_M$. Hence B_M is $d_{\max}$-closed.

Solution to Exercise 1.10

Suppose that (f_n) is a $d_{\max}$-convergent sequence of functions in $C[0,1]$ with $f_n \in C$ for each n. We must show that the limit of the sequence, f, say, also lies in C. That is, we must show $|f(a)| \leq M$.

For each $n \in \mathbb{N}$, we can use the Triangle Inequality to find an upper bound for $|f(a)|$ involving f_n as follows:

$$\begin{aligned} |f(a)| = |(f(a) - f_n(a)) + f_n(a)| &\leq |f(a) - f_n(a)| + |f_n(a)| \\ &\leq |f(a) - f_n(a)| + M, \end{aligned}$$

since $f_n \in C$. But $|f(a) - f_n(a)| \leq d_{\max}(f_n, f)$ and so

$$|f(a)| \leq d_{\max}(f_n, f) + M \text{ for each } n \in \mathbb{N}.$$

Yet $(d_{\max}(f_n, f))$ is a null sequence and so we conclude that

$$|f(a)| \leq M,$$

implying that $f \in C$. Thus C is $d_{\max}$-closed, as required.

Solution to Exercise 1.11

We saw in Lemma 1.2 that the singleton sets $\{x_i\}$ are d-closed for each $i = 1, 2, \ldots, n$ and so we can use the remark following Proposition 1.14 to conclude that their union,

$$\{x_1, x_2, \ldots, x_n\} = \bigcup_{i=1}^{n} \{x_i\},$$

is also d-closed.

Solution to Exercise 1.12

There are many possible solutions to this question and your answer may differ from ours.

Let $A_n = \{\frac{1}{n}\}$ for $n \in \mathbb{N}$. Then each set A_n is a singleton and so is closed by Lemma 1.2. However, by considering the sequence (a_n) given by $a_n = \frac{1}{n}$, which has limit 0, we see that $A = \cup_{n=1}^{\infty} A_n = \{\frac{1}{n} : n \in \mathbb{N}\}$ is not a closed set, since (a_n) is an example of a convergent sequence with terms lying in A whose limit is not in A.

Solution to Exercise 1.13

(a) $K_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ and $K_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$.

(b) Each K_n is a finite union of closed intervals (each of which is closed with respect to the Euclidean metric) and so, by Proposition 1.14, K_n is itself closed for the Euclidean metric.

Solution to Exercise 1.14

We observe that the sphere (in X) can be written as the intersection of two closed sets:

$$\begin{aligned} S_d(a,r) &= \{x \in X : d(a,x) \leq r\} \cap \{x \in X : d(a,x) \geq r\} \\ &= B_d[a,r] \cap \{x : d(a,x) \geq r\} \\ &= B_d[a,r] \cap (B_d(a,r))^c, \end{aligned}$$

and so, by Proposition 1.17, $S_d(a,r)$ is closed.

We showed in Worked Exercise 1.6 that $B_d(a,r)^c$ is d-closed.

Solution to Exercise 1.15

(a) The argument is similar to that given in the solution to Exercise 1.9.

Suppose that (f_n) is a $d_{\max}$-convergent sequence of functions in $C[0,1]$ with $f_n \in F_M$ for each n. We must show that the limit of the sequence – f, say – also lies in F_M. That is, we must show $f(\frac{1}{M}) = 0$.

For each $n \in \mathbb{N}$ we can find an upper bound for $|f(\frac{1}{M})|$ involving f_n as follows:

$$\begin{aligned} \left|f(\tfrac{1}{M})\right| &= \left|f(\tfrac{1}{M}) - 0\right| \\ &= \left|f(\tfrac{1}{M}) - f_n(\tfrac{1}{M})\right| \\ &\leq d_{\max}(f, f_n). \end{aligned}$$

But $(d_{\max}(f_n, f))$ is a null sequence and so we conclude (by the Squeeze Rule) that

$$|f(\tfrac{1}{M})| = 0,$$

implying that $f \in F_M$. We conclude that F_M is $d_{\max}$-closed.

(b) The function $f\colon [0,1] \to \mathbb{R}$ given by $f(x) = 0$ for each $x \in [0,1]$ lies in the intersection. In fact, any function $g \in C[0,1]$ that is zero at each of the points $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ will lie in the intersection of the sets. (Notice that, because functions in $C[0,1]$ are continuous, if a function $g \in C[0,1]$ is zero at each of $1, \frac{1}{2}, \frac{1}{3}, \ldots$, then it must also be zero at zero: $g(0) = 0$.)

(c) Since, by part (a), each of the sets F_M is $d_{\max}$-closed, we conclude from Proposition 1.17 that their intersection is also $d_{\max}$-closed.

Solution to Exercise 2.1

To show that (a,∞) is open, we show that its complement in $\mathbb{R}$ is closed: $\mathbb{R} - (a,\infty) = (-\infty, a]$ and we saw in Worked Exercise 1.4 that this set is closed for the Euclidean metric.

The interval $[a,\infty)$ is not open for the Euclidean metric because the complement of $[a,\infty)$ (in $\mathbb{R}$) is $(-\infty, a)$ and this is not a closed set: the sequence (a_n) given by $a_n = a - \frac{1}{n}$ lies inside $(-\infty, a)$ for each n but the limit does not.

Solution to Exercise 2.2

The complement of the interval $(a, b]$ in $\mathbb{R}$ is $(-\infty, a] \cup (b, \infty)$ and by considering the sequence $x_n = b + \frac{1}{n}$ for $n \in \mathbb{N}$, we see that the complement is not closed for the Euclidean metric.

Similarly, we see that the complement of the interval $[a, b]$ in $\mathbb{R}$ is $(-\infty, a) \cup (b, \infty)$ and consideration of the same sequence shows that this is not a closed set for the Euclidean metric.

Solution to Exercise 2.3

If we look at the complement of C_M in $C[0, 1]$, we see that

$$C[0, 1] - C_M = \{f \in C[0, 1] : |f(x)| \leq M \text{ for each } x \in [0, 1]\}$$

and this is the set that we determined to be closed (for $d_{\max}$) in Worked Exercise 1.12. Hence C_M is open.

Solution to Exercise 2.4

Let $f_0 \in C[0, 1]$ be the function defined by

$$f_0(x) = 0 \quad \text{for each } x \in [0, 1].$$

Then $f_0 \in S$. We show that there is no $r > 0$ so that $B_{d_{\max}}(f_0, r)$ lies inside S.

So let $r > 0$. We must find a function g with $d_{\max}(f_0, g) < r$, but $g(0) \neq 0$ and so $g \notin S$.

Let $g \in C[0, 1]$ be the function given by

$$g(x) = \tfrac{1}{2}r \quad \text{for each } x \in [0, 1].$$

Then $d_{\max}(f_0, g) = \frac{1}{2}r < r$, and so $g \in B_{d_{\max}}(f_0, r)$. But $g(0) = \frac{1}{2}r \neq 0$ and so $g \notin S$. Thus, S is not open.

Solution to Exercise 2.5

Combining the results of Worked Exercise 2.4 and Exercise 2.3 of Chapter 15, we see that for $\mathbf{a} \in \mathbf{C}$ and $n \in \mathbb{N}$,

$$B_{d_{\mathbf{C}}}(\mathbf{a}, 2^{-n}) = B_{d_{\mathbf{C}}}[\mathbf{a}, 2^{-(n+1)}] = (a_1, a_2, \ldots, a_n, c_1, c_2, \ldots) : (c_i) \in \mathbf{C}.$$

Thus open balls and closed balls coincide in $\mathbf{C}$, and so it follows from Proposition 2.3 that the $d_{\mathbf{C}}$-closed balls are $d_{\mathbf{C}}$-open!

Solution to Exercise 2.6

Since (a_n) does not converge to a, there is $\varepsilon > 0$ such that for each $N \in \mathbb{N}$ there is an $n > N$ with $a_n \notin B_d(a, \varepsilon)$. Let (a_{m_n}) be the subsequence of (a_n) formed as follows:

- Let m_1 be the first integer such that $a_{m_1} \notin B_d(a, \varepsilon)$.
- Let $m_2 > m_1$ be the first integer strictly larger than m_1 such that $a_{m_2} \notin B_d(a, \varepsilon)$.
- For $n > 2$, let m_n be the first integer strictly larger than m_{n-1} for which $a_{m_n} \notin B_d(a, \varepsilon)$.

Then (a_{m_n}) is a subsequence with the required property.

Solution to Exercise 2.7

$A_1 = \{2, 4, 6, 8, 10, 12\}$, $A_2 = \{3, 6, 9, 12\}$;
$A_1 \cap A_2 = \{6, 12\}$, $A_1 \cup A_2 = \{2, 3, 4, 6, 8, 9, 10, 12\}$.
Thus, $A_1^c = \{1, 3, 5, 7, 9, 11\}$; $A_2^c = \{1, 2, 4, 5, 7, 8, 10, 11\}$;
$(A_1 \cup A_2)^c = \{1, 5, 7, 11\} = A_1^c \cap A_2^c$;
$(A_1 \cap A_2)^c = \{1, 2, 3, 4, 5, 7, 8, 9, 10, 11\} = A_1^c \cup A_2^c$.

Solution to Exercise 2.8

First Law: $(A_1 \cup A_2 \cup A_3)^c = A_1^c \cap A_2^c \cap A_3^c$. In other words,

$$X - (A_1 \cup A_2 \cup A_3) = (X - A_1) \cap (X - A_2) \cap (X - A_3).$$

Second Law: $(A_1 \cap A_2 \cap A_3)^c = A_1^c \cup A_2^c \cup A_3^c$. In other words,

$$X - (A_1 \cap A_2 \cap A_3) = (X - A_1) \cup (X - A_2) \cup (X - A_3).$$

Solution to Exercise 2.9

Let (X, d) be a metric space, and let U be the union of a collection of d-open sets $\{U_i : i \in I\}$.

We must show that U is d-open: that is, $X - U$ is d-closed. We use De Morgan's First Law to find

$$\begin{aligned} X - U &= X - \left(\bigcup_{i \in I} U_i \right) \\ &= \bigcap_{i \in I} (X - U_i). \end{aligned}$$

But for each $i \in I$, $X - U_i$ is d-closed since U_i is d-open and so, by Proposition 1.17, $\bigcap_{i \in I}(X - U_i)$ is d-closed. That is, $X - U$ is d-closed and so U is d-open, as required.

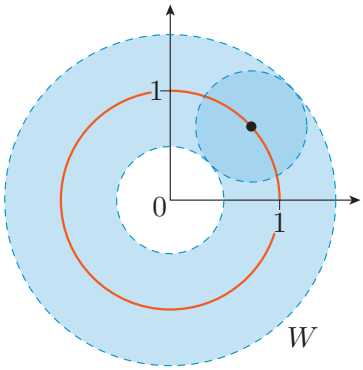
Solution to Exercise 2.10

(a) For each $(x, y) \in U$, $y - 1 > 0$. Let $r_{x,y} = y - 1$ and consider the open balls $B_d((x, y), r_{x,y})$. U is the union of these infinitely many open balls, that is,

$$U = \bigcup_{(x,y)\in U} B_d((x, y), r_{x,y})$$

and so U is the union of (an infinite collection of) open sets. Therefore, by Theorem 2.15, U is d-open.

(b) $V = B_d((1, 1), 1) \cap B_d((2, 2), 1)$; that is, V is the intersection of two open balls. Therefore, by Theorem 2.13, V is open.



(c) The point (x, y) belongs to W if, and only if, it is at a distance less than $\frac{1}{2}$ from some point (a, b) on the unit circle. That is, W is the union of an infinite collection of open balls of radius $\frac{1}{2}$, as shown in the figure:

$$W = \bigcup_{(a,b):a^2+b^2=1} B_d((a, b), \tfrac{1}{2}).$$

Hence by Theorem 2.15, W is open.

Solution to Exercise 3.1

Since we do not yet have any defined method for associating an open or closed set with a given set, you can view your answers as being correct provided that they are of the correct kind (open or closed), regardless of whether they match ours!

(a) Proceeding by analogy with the interval $(0, 1]$, we would associate the closed interval $[0, 1]$ and the open interval $(0, 1)$.

(b) The interval $[0, 1]$ is closed and so the natural closed set to associate with it is itself, $[0, 1]$.

For an open set: there are many open sets that contain the interval $[0, 1]$ – for example any interval of the form (a, b), where $a < 0$ and $b > 1$ – and there appears to be no canonical way to choose a particular one. However, if we look at open sets that are contained in $[0, 1]$, then we discover that there appears to be a 'largest' such set, namely the open interval $(0, 1)$ – any other open set inside $[0, 1]$ lies inside this open interval, and so $(0, 1)$ is a natural open set to associate with $[0, 1]$.

(c) The interval $(0, 1)$ is open and so the natural open set to associate with it is itself, $(0, 1)$.

For a closed set: there are many closed sets that lie inside the interval $(0, 1)$ – for example any interval of the form $[a, b]$, where $0 < a \leq b < 1$ – and there appears to be no canonical way to choose one. However, if we look at closed sets that contain $(0, 1)$, then we discover that there appears to be a 'smallest' such set, namely the closed interval $[0, 1]$, and so this is a natural closed set to associate with $[0, 1]$.

(d) This is a set consisting of a single point and so is a closed set itself. Thus the naturally associated closed set is $\{0\}$.

To associate an open set is a little trickier: there are lots of open sets that contain $\{0\}$ – for example, any interval of the form (a, b) where $a < 0$ and $b > 0$ – and none of them seem more natural than any other. Regarding open subsets of $\{0\}$, our choice here is limited: the only open subset of $\{0\}$ is the empty set, $\emptyset$, and we choose this as our natural open set associated with $\{0\}$.

Solution to Exercise 3.2

The empty set is a closed set (that contains itself) and so

$$\emptyset \subseteq \mathrm{Cl}_{(X,d)}(\emptyset) \subseteq \emptyset.$$

Hence $\mathrm{Cl}_{(X,d)}(\emptyset) = \emptyset$.

Similarly X is a closed set (that contains itself) and so

$$X \subseteq \mathrm{Cl}_{(X,d)}(X) \subseteq X.$$

Hence $\mathrm{Cl}_{(X,d)}(X) = X$.

Solution to Exercise 3.3

We know that $(0, 1)$ is not closed for the Euclidean metric and that $(0, 1) \subseteq \mathrm{Cl}((0, 1))$. We also know that $[0, 1]$ is a closed set that contains $(0, 1)$ and so

$$(0, 1) \subseteq \mathrm{Cl}((0, 1)) \subseteq [0, 1].$$

It remains only to show that there are no proper subsets of $[0, 1]$ that contain $(0, 1)$ and are closed. By considering the sequence $\left(\frac{1}{n+1}\right)_{n\in\mathbb{N}}$ we see that $(0, 1]$ is not closed and a similar argument also shows that $[0, 1)$ is not closed. Thus there are no closed sets F lying between $(0, 1)$ and $[0, 1]$. Hence $\mathrm{Cl}((0, 1)) = [0, 1]$.

Solution to Exercise 3.4

Following the method of the solution to Exercise 1.15, we observe that

$$\begin{aligned} A &= \{f \in C[0, 1] : f(x) = 0 \text{ for } x \in I\} \\ &= \bigcap_{x\in I} \{f \in C[0, 1] : f(x) = 0\}. \end{aligned}$$

This is an intersection of closed sets. It follows from Proposition 1.17 that A is closed, and so, by Theorem 3.3, $\mathrm{Cl}(A) = A$.

Solution to Exercise 3.5

It is certainly the case that both $A \subseteq \mathrm{Cl}(A)$ and $B \subseteq \mathrm{Cl}(B)$ and so

$$A \cap B \subseteq \mathrm{Cl}(A) \cap \mathrm{Cl}(B),$$

but $\mathrm{Cl}(A) \cap \mathrm{Cl}(B)$ is a closed set and so

$$\mathrm{Cl}(A \cap B) \subseteq \mathrm{Cl}(A) \cap \mathrm{Cl}(B),$$

as required.

Solution to Exercise 3.6

By the definition of complements, $\mathbb{Q} \cap \mathbb{Q}^c = \emptyset$ and, by the definition of closure, $\mathrm{Cl}(\emptyset) = \emptyset$. Thus

$$\mathrm{Cl}(\mathbb{Q} \cap \mathbb{Q}^c) = \emptyset.$$

However, by Worked Exercise 3.9, $\mathrm{Cl}(\mathbb{Q}) = \mathrm{Cl}(\mathbb{Q}^c) = \mathbb{R}$, and so

$$\mathrm{Cl}(\mathbb{Q}) \cap \mathrm{Cl}(\mathbb{Q}^c) = \mathbb{R}.$$

Therefore, for $A = \mathbb{Q}$ and $B = \mathbb{Q}^c$, $\mathrm{Cl}(A \cap B)$ is a proper subset of $\mathrm{Cl}(A) \cap \mathrm{Cl}(B)$.

Solution to Exercise 3.7

Since (a_n) is an increasing sequence that is bounded above, it has a limit, a, say. We know that $A \subseteq \mathrm{Cl}(A)$, so it only remains to determine which points in A^c belong to $\mathrm{Cl}(A)$.

If $x \in A^c$ and $x \neq a$, then either $x < a_1$, or $x > a$, or (since (a_n) is strictly increasing) there is an $n \in \mathbb{N}$ for which

$$a_n < x < a_{n+1}.$$

In each case, we can find $r > 0$ such that $B(x, r) \cap A = (x - r, x + r) \cap A = \emptyset$, and so x is not a closure point of A.

If $x = a$, then for each $r > 0$, the definition of convergence of a sequence means that there is an $N \in \mathbb{N}$ such that $a_n \in B(a, r)$ for each $n > N$. Hence $B(a, r) \cap A = (a - r, a + r) \cap A \neq \emptyset$. Thus a is a closure point of A.

Therefore

$$\mathrm{Cl}(A) = A \cup \{a\}.$$

Solution to Exercise 3.8

Let $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and let $r > 0$ be given.

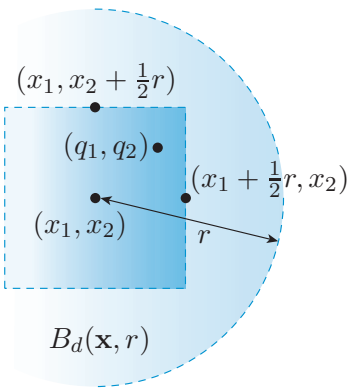
The density property of the reals means that between any two distinct real numbers there is a rational number, and so $(x_1 - \frac{1}{2}r, x_1 + \frac{1}{2}r) \cap \mathbb{Q} \neq \emptyset$ and $(x_2 - \frac{1}{2}r, x_2 + \frac{1}{2}r) \cap \mathbb{Q} \neq \emptyset$. Hence

$$((x_1 - \tfrac{1}{2}r, x_1 + \tfrac{1}{2}r) \times (x_2 - \tfrac{1}{2}r, x_2 + \tfrac{1}{2}r)) \cap (\mathbb{Q} \times \mathbb{Q}) \neq \emptyset.$$

Moreover,

$$(x_1 - \tfrac{1}{2}r, x_1 + \tfrac{1}{2}r) \times (x_2 - \tfrac{1}{2}r, x_2 + \tfrac{1}{2}r) \subseteq B_d(\mathbf{x}, r).$$

Hence $B_d(\mathbf{x}, r)$ intersects X, as shown in the figure.



Thus each open ball around $\mathbf{x}$ meets X implying, by Theorem 3.8, that $\mathbf{x} \in \mathrm{Cl}(X)$. Thus $\mathbb{R}^2 \subseteq \mathrm{Cl}(X) \subseteq \mathbb{R}^2$, as required.

Solution to Exercise 4.1

Since $\emptyset$ is an open set, $\emptyset \subseteq \text{Int}(\emptyset)$ but we also know that $\text{Int}(\emptyset) \subseteq \emptyset$ and so $\text{Int}(\emptyset) = \emptyset$.

Similarly, X is an open set and so

$$X \subseteq \text{Int}(X) \subseteq X,$$

giving $\text{Int}(X) = X$.

Solution to Exercise 4.2

Suppose that $A \subseteq B$; then $\text{Int}(A)$ is an open subset of X that is contained in A and so it is also an open set contained in B. Hence, by the definition of $\text{Int}(B)$, $\text{Int}(A) \subseteq \text{Int}(B)$.

Solution to Exercise 4.3

We have $A \cup B = [0, 2]$ and so $\text{Int}(A \cup B) = (0, 2)$. Thus

$$\begin{aligned}\text{Int}(A) \cup \text{Int}(B) &= (0, 1) \cup (1, 2) \\ &= \text{Int}(A \cup B) - \{1\}.\end{aligned}$$

Solution to Exercise 4.4

Let $q \in \mathbb{Q}$; then by the density property of the reals, for each $r > 0$, the ball $B_d(q, r) \cap \mathbb{Q}^c \neq \emptyset$ and so, by Theorem 4.4, $q \notin \text{Int}(\mathbb{Q})$. Since $\text{Int}(\mathbb{Q}) \subseteq \mathbb{Q}$, we conclude that $\text{Int}(\mathbb{Q}) = \emptyset$.

Solution to Exercise 4.5

A set A is open if, and only if, A^c is closed – which, by Theorem 3.3, holds if, and only if, $A^c = \text{Cl}(A^c)$. But, by Theorem 4.6, $[\text{Cl}(A^c)]^c = \text{Int}(A)$ and so A is open if, and only if,

$$A = (A^c)^c = [\text{Cl}(A^c)]^c = \text{Int}(A).$$

Solution to Exercise 5.1

By Theorem 5.2,

$$\begin{aligned}\text{Bd}(A^c) &= \text{Cl}(A^c) \cap \text{Cl}((A^c)^c) \\ &= \text{Cl}(A^c) \cap \text{Cl}(A) = \text{Bd}(A).\end{aligned}$$

Solution to Exercise 5.2

(a) By Worked Exercise 3.9, $\text{Cl}(\mathbb{Q}) = \text{Cl}(\mathbb{Q}^c) = \mathbb{R}$. It follows from Theorem 4.6 that

$$\text{Int}(\mathbb{Q}) = [\text{Cl}(\mathbb{Q}^c)]^c = \mathbb{R}^c = \emptyset$$

and

$$\text{Int}(\mathbb{Q}^c) = [\text{Cl}(\mathbb{Q})]^c = \mathbb{R}^c = \emptyset.$$

Hence by Theorem 5.5,

$$\text{Bd}(\mathbb{Q}) = (\text{Int}(\mathbb{Q}^c) \cup \text{Int}(\mathbb{Q}))^c = \emptyset^c = \mathbb{R}.$$

(b) For $\mathbb{Q}^c$, similar arguments can be used to obtain $\text{Bd}(\mathbb{Q}^c) = \mathbb{R}$.

Solution to Exercise 5.3

For the discrete metric, every subset of X is both open and closed. So A and A^c are open and, by Exercise 4.5, $\mathrm{Int}(A) = A$ and $\mathrm{Int}(A^c) = A^c$.

By Theorem 5.5,

$$\begin{aligned}\mathrm{Bd}(A) &= (\mathrm{Int}(A^c) \cup \mathrm{Int}(A))^c \\ &= (A^c \cup A)^c = X^c = \emptyset.\end{aligned}$$

Solution to Exercise 5.4

We note that

$$\begin{aligned}x \in (E \cup F) \cap G &\text{ if, and only if, } x \in (E \cup F) \text{ and } x \in G \\ &\text{if, and only if, } (x \in E \text{ or } x \in F) \text{ and } x \in G \\ &\text{if, and only if, } (x \in E \text{ and } x \in G) \text{ or } (x \in F \text{ and } x \in G) \\ &\text{if, and only if, } (x \in E \cap G) \text{ or } (x \in F \cap G) \\ &\text{if, and only if, } x \in (E \cap G) \cup (F \cap G),\end{aligned}$$

as required.

Solution to Exercise 5.5

It follows from the result of Exercise 5.1 that

$$\mathrm{Bd}(A \cap B) = \mathrm{Bd}((A \cap B)^c).$$

Using De Morgan's Second Law, this gives

$$\mathrm{Bd}(A \cap B) = \mathrm{Bd}(A^c \cup B^c).$$

It now follows from Theorem 5.6 and the result of Exercise 5.1 that

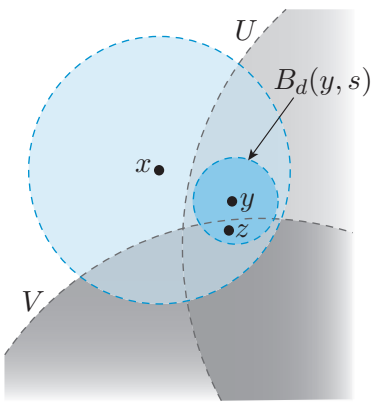
$$\mathrm{Bd}(A \cap B) \subseteq \mathrm{Bd}(A^c) \cup \mathrm{Bd}(B^c) = \mathrm{Bd}(A) \cup \mathrm{Bd}(B).$$

Solution to Exercise 6.1

Since $B \subseteq A$, we have $\mathrm{Cl}(B) \subseteq \mathrm{Cl}(A)$ and so

$$\mathrm{Int}(\mathrm{Cl}(B)) \subseteq \mathrm{Int}(\mathrm{Cl}(A)) = \emptyset.$$

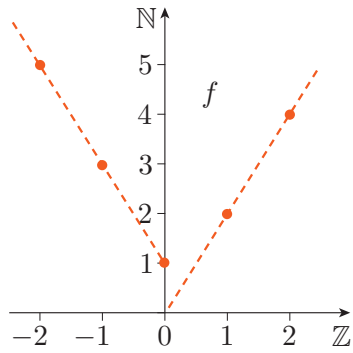
Solution to Exercise 6.2



Since U and V are open, so is $U \cap V$. So we need only show that $U \cap V$ is dense in X – that is, $\mathrm{Cl}(U \cap V) = X$. We do this by using the characterisation of closed sets in terms of closure points, Theorem 3.8, and show that for each $x \in X$ and $r > 0$, $B_d(x, r)$ meets $U \cap V$.

Let $x \in X$ and let $r > 0$ be given. Since U is dense in X, $B_d(x, r) \cap U \neq \emptyset$ and we can choose $y \in B_d(x, r) \cap U$. Moreover, since $B_d(x, r) \cap U$ is open, we can find $s > 0$ so that $B_d(y, s) \subseteq B_d(x, r) \cap U$. But V is dense in X and so $\mathrm{Cl}(V) = X$. In particular, $B_d(y, s) \cap V \neq \emptyset$ and we can choose $z \in B_d(y, s) \cap V$ (as shown). But then $z \in U \cap V \cap B_d(x, r)$ and so $B_d(x, r)$ meets $U \cap V$. We deduce that $x \in \mathrm{Cl}(U \cap V)$ and so $U \cap V$ is dense in X as required.

Solution to Exercise 6.3



The listing of the elements of $\mathbb{Z}$ as

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

leads to the one–one map from $\mathbb{Z}$ to $\mathbb{N}$ defined by

$$f(n) = \begin{cases} 1, & \text{for } n = 0, \\ 2n, & \text{for } n > 0, \\ -2n + 1, & \text{for } n < 0, \end{cases}$$

as shown in the figure. (There are other possible maps.) So $\mathbb{Z}$ is countable.

Solution to Exercise 6.4

Suppose that (a, b) is countable, so that there is a one–one map $f\colon (a, b) \to \mathbb{N}$. Consider the map $g\colon (0, 1) \to (a, b)$ given by

$$g(x) = a + (b - a)x.$$

This map is one–one since if $g(x) = g(y)$, so that $a + (b - a)x = a + (b - a)y$, then $(b - a)x = (b - a)y$; hence, since $a \neq b$, $x = y$. Therefore the map $f \circ g$ is a one–one map from $(0, 1)$ to $\mathbb{N}$. But this is impossible, since $(0, 1)$ is uncountable (by Theorem 6.7). Hence (a, b) must be uncountable.

Solution to Exercise 7.1

Observe that $x \in f^{-1}(Y - A)$ if, and only if, $f(x) \in Y - A$; and $f(x) \in Y - A$ if, and only if, $f(x) \in Y$ and $f(x) \notin A$, which holds if, and only if, $x \in f^{-1}(Y)$ and $x \notin f^{-1}(A)$ – in other words if, and only if, $x \in X - f^{-1}(A)$.

Hence $x \in f^{-1}(Y - A)$ if, and only if, $x \in X - f^{-1}(A)$ and so $f^{-1}(Y - A) = X - f^{-1}(A)$ as required.

Index

$(C[0,1], d_{\max})$
  closed sets 188
  open sets 196
$\emptyset$ 7

absolute value 6

$B_d(a,r)$, the open ball of radius r with centre a 94
$B_d[a,r]$, the closed ball of radius r with centre a 94
balls
  Cantor metric 130
  closed 94
  Hamming distance 125
  max metric 156
  open 94
basic continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 31
basic null sequences 14
$\mathrm{Bd}_{(X,d)}(A)$, the boundary of A in (X,d) 220
boundary of a set 220
boundary points of a set 221
bounded function 34, 147
bounded set 34
Boundedness Theorem 34

C, Cantor space 126
$C[0,1]$, the continuous functions from $[0,1]$ to $\mathbb{R}$ 152
$C[a,b]$, the continuous functions from $[a,b]$ to $\mathbb{R}$ 152
Cantor distance 128
Cantor metric 129
  balls 130
  continuity 131
Cantor set
  middle-third 193
Cantor space 126
  closed sets 188
Cartesian product 66
Cauchy–Schwarz Inequality 83, 85
centre 94
$\mathrm{Cl}(A)_{(X,d)}$, the closure of A in (X,d) 208
closed rectangle 186
closed ball 94
closed bounded interval 32
closed set
  $(C[0,1], d_{\max})$ 188
  Cantor space 188
  closed intervals in $\mathbb{R}$ 182
  discrete metric 187
  in the plane 185
  intersection 192
  spheres 192
  union 190
closure of a set 208
closure point 211
codomain 5
combination rules
  for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 29
  for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ 47
  for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ 79
  for convergent sequences in $\mathbb{R}$ 19
  for convergent sequences in $\mathbb{R}^n$ 75
  for null sequences 15
complement of a set 183
component sequence 142
composition of functions 30
Composition Rule
  for continuous functions between metric spaces 103
  for continuous functions on $\mathbb{R}$ 30
  for continuous functions on $\mathbb{R}^2$ 48
  for continuous functions on $\mathbb{R}^n$ 80
constant sequence 11
continuity at a point
  for Euclidean spaces 76
  for metric spaces 102
  in the plane 43
  on the real line 24
continuity for metric spaces 102
continuity on set
  in $\mathbb{R}$ 24
  in $\mathbb{R}^2$ 43
  in $\mathbb{R}^n$ 76
continuous function
  between metric spaces 102
  from $\mathbb{R}$ to $\mathbb{R}$ 24
  from $\mathbb{R}^2$ to $\mathbb{R}$ 43
  from $\mathbb{R}^n$ to $\mathbb{R}^m$ 76
  sequential definition 24
continuous functions between metric spaces
  Composition Rule 103
continuous functions from $\mathbb{R}$ to $\mathbb{R}$
  combination rules 29
  Composition Rule 30
  Restriction Rule 30
continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$
  combination rules 47
  Composition Rule 48
  Restriction Rule 48
continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$
  combination rules 79
  Composition Rule 80
  Restriction Rule 79
convergence
  pointwise 144
  product space 142
  uniform 147

convergent sequence
in $\mathbb{R}$ 16
in $\mathbb{R}^2$ 40
in $\mathbb{R}^n$ 72
in a metric space 99
convergent sequences
combination rules 19, 75
coordinate sequence 73, 142
countable set 227
countably infinite set 227

d-closed set 180
d-dense set 225
d_H, the Hamming distance 125
$d_{\max}$, the max metric on $C[0,1]$ 153
d-open set 194
De Morgan's Laws 202
d-closed ball 94
definition
d-convergent sequence 99
d-divergent sequence 99
Cantor distance 128
Cantor metric 129
closed ball 94
continuous function between metric spaces 102
convergent sequence in a metric space 99
convergent sequences in $\mathbb{R}$ 16
convergent sequences in $\mathbb{R}^2$ 40
convergent sequences in $\mathbb{R}^n$ 72
discontinuous function between metric spaces 102
divergent sequences in $\mathbb{R}$ 16
divergent sequences in $\mathbb{R}^2$ 40
divergent sequences in $\mathbb{R}^n$ 72
eventually constant 100
Lipschitz function 132
metric space 88
null sequence 12
null sequence in $\mathbb{R}^n$ 72
open ball 94
sequence in a metric space 98
sphere 94
dense set 225
discontinuity
essential 29
removable 29
discontinuity on set
in $\mathbb{R}$ 24
in $\mathbb{R}^2$ 43
in $\mathbb{R}^n$ 76
discontinuous function
between metric spaces 102
from $\mathbb{R}$ to $\mathbb{R}$ 24
from $\mathbb{R}^2$ to $\mathbb{R}$ 43
from $\mathbb{R}^n$ to $\mathbb{R}^m$ 76
discrete metric 89
boundary 223
closed sets 187
interior 223
open sets 196
distance
in $\mathbb{R}$ 9
in a metric space 88
on $\mathbb{R}^n$ (Euclidean) 68
on $\mathbb{R}^2$ (Euclidean) 39
distance function 68, 87
d-open ball 94
d-sphere 94
divergence 99
divergent sequence
in $\mathbb{R}$ 16
in $\mathbb{R}^2$ 40
in $\mathbb{R}^n$ 72
in a metric space 99
domain 5

e_1, e_2, e_∞ 121
empty set 7
closed 181
open 195
equivalent metrics 136
Euclidean metric 88
Euclidean distance 49
non-negativity property 69
on $\mathbb{R}^2$ 39
on $\mathbb{R}^n$ 68
symmetry property 69
Triangle Inequality 50, 69
Euclidean distance function 87
Euclidean n-space 88
even subsequence 21
eventually constant 100
eventually zero 129
Extreme Value Theorem 35

finite set 227
function 5
bounded 34, 147
codomain 5
continuous from $\mathbb{R}$ to $\mathbb{R}$ 24
continuous from $\mathbb{R}^2$ to $\mathbb{R}$ 43
continuous from $\mathbb{R}^n$ to $\mathbb{R}^m$ 76
(d,e)-continuous 102
discontinuous from $\mathbb{R}$ to $\mathbb{R}$ 24
discontinuous from $\mathbb{R}^2$ to $\mathbb{R}$ 43
discontinuous from $\mathbb{R}^n$ to $\mathbb{R}^m$ 76
(d,e)-discontinuous 102
domain 5
image of a point 5
image of a set 5

Lipschitz 132
preimage of a set 5
restriction of 29, 120
rule 5
functions
composition of 30
functions from $\mathbb{R}$ to $\mathbb{R}$
basic continuous 31

Hamming distance 125
balls 125

image
of a point 5
of a set 5
index set 192
induced metric 119
induced subspace 119
inequality
solution set 7, 9
infinite set 227
integer part 77
integers 4
integral metric 155
interior
discrete metric 223
of a set 216
interior point 219
Intermediate Value Theorem 32
intersection
of closed sets 192
of open sets 202
of sets 192
interval
closed 4
open 4
inverse image
of a set 5
irrational numbers 4

limit of a sequence 16, 40, 72, 99
pointwise 144
uniform 147
Lipschitz constant 132
Lipschitz function 132
listable set 228
lower bound 20

map 4
max metric
continuity 157
on $C[0,1]$ 153
on $\mathbb{R}^2$ 123
open and closed balls 156
open sets 196
metric 87, 88
Cantor 129
discrete 89
equivalent 136
Euclidean 88
Hamming 125
induced 119
integral 155
max 123, 153
mixed (on the plane) 93
product 121
pull-back 116
$\tan^{-1}$ 118
subspace 119
taxicab 91, 96
transferring 115
metric space 88
boundary 220
boundary points 221
closed set 180
closure point 211
dense set 225
interior point 219
nowhere dense 225
open set 194
points 88
Reverse Triangle Inequality 89
metric subspace 119
metrisable topology 233
middle-third Cantor set 193
mixed metric in the plane 93
modulus 6
distance 6
Monotone Convergence Theorem for sequences in $\mathbb{R}$ 20
monotonic decreasing and increasing sequences in $\mathbb{R}$ 11
Multiple Rule
for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 29
for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ 47
for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ 79
for convergent sequences 19, 75
for null sequences 15

natural numbers 4
nowhere dense 225
metric space 225
nth term of sequence 11, 37
null sequence
in $\mathbb{R}$ 12
in $\mathbb{R}^n$ 72
null sequences
basic 14
combination rules 15
Squeeze Rule 15

odd subsequence 21
open ball 94
open balls are open 195
open rectangle 186
open sets
    characterisation via open balls 198
    definition 194
    discrete metric 196
    in $(C[0,1], d_{\max})$ 196
    intersection 202
    max metric 196
    open balls 195
    topology 233
    union 204
origin 36, 67

plane
    closed set 185
    open set 195
point in a metric space 88
pointwise convergence 144
pointwise limit 144
Power Rule
    for null sequences 15
preimage
    of a set 5
product metric 121
Product Rule
    for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 29
    for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ 47
    for convergent sequences 19
    for null sequences 15
product sequence 142
product set 66
projection function 139
    continuity 139
    from $\mathbb{R}^2$ to $\mathbb{R}$ 43
    from $\mathbb{R}^n$ to $\mathbb{R}$ 80
properties of Euclidean distance 49
    non-negativity 50
    symmetry 50
    Triangle Inequality in $\mathbb{R}^2$ 50
pull-back of d by f 116

Quotient Rule
    for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 29
    for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ 47
    for convergent sequences 19

radius 94
rational numbers 4
real numbers 4
real sequence 11
real-valued function 152
restriction of a function 29, 120
Restriction Rule
    for continuous functions between metric spaces 120
    for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 30
    for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ 48
    for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ 79
Reverse Triangle Inequality
    for metrics 89
    in $\mathbb{R}$ 10
    in $\mathbb{R}^n$ 70

$S_d(a,r)$, the sphere of radius r with centre a 94
sequence
    constant 11
    eventually constant 100
    eventually zero 129
    in A 11
    nth term 11
sequence in a metric space 98
    convergence 99
    divergence 99
    limit 99
sequence space
    Cantor 126
    eventually zero sequences 129
sequences in $\mathbb{R}$ 11
    basic null 14
    convergence 16
    divergence 16
    Monotone Convergence Theorem 20
    monotonic decreasing 11
    monotonic increasing 11
    null 12
sequences in $\mathbb{R}^2$ 37
    convergence 40
    divergence 40
sequences in $\mathbb{R}^n$ 71
    convergence 72
    divergence 72
sequential definition of continuity 24
set
    boundary point 221
    closed 180
    closure point 211
    complement 183
    countable 227
    d-closed 180
    dense 225
    finite 227
    index 192
    infinite 227
    interior point 219
    listable 228
    nowhere dense 225
    open 194

uncountable 227
uncountably infinite 227
set of symbols 125
words of length n 125
sets
De Morgan's Laws 202
intersection 192
union 203
shift map 131
single point is closed 181
sphere 94
spheres are closed sets 192
Squeeze Rule for null sequences 15
subsequence 21
even 21
odd 21
subspace
induced metric 119
metric 119
Sum Rule
for continuous functions from $\mathbb{R}$ to $\mathbb{R}$ 29
for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ 47
for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ 79
for convergent sequences 19, 75
for null sequences 15
surface 36

$\tan^{-1}$ metric 118
taxicab metric 91
open balls 95
topological space 233
topology 233
metrisable 233
open sets 233
totally disconnected set 193
transferring a metric 115
Triangle Inequality
condition for metrics 88
in $\mathbb{R}$ 9
in $\mathbb{R}^n$ 69, 85, 87

uncountable set 227
uniform convergence 147
Uniform Convergence Theorem 161
uniform limit 147
union
of closed sets 190
of open sets 204
of sets 203
unit closed ball 94
unit open ball 94
unit sphere 94
upper bound 20

whole space
closed 181
open 195